nature

www.nature.com/nature

Vol 446 | Issue no. 7134 | 22 March 2007

Millennium development holes

The political commitment to helping the developing world is failing to deliver on its promises. The problem is made worse by the questionable evaluation of progress.

In 2000, 189 world leaders committed to eight Millennium Development Goals (MDGs), ranging from halving extreme poverty and hunger, and rolling back killer diseases such as AIDS and malaria, to providing universal primary education. The deadline of 2015 to achieve all these ambitious goals is now rapidly approaching.

As a rallying cry that has pushed development up the international political agenda, the goals have been an indisputable success. They have also, for better or worse, conferred power and legitimacy on interests within the international aid machine, in particular the United Nations (UN) and the World Bank.

But the goals are ultimately political promises, and as such they are fickle. As the economist Jeffrey Sachs pointed out at last week's 'BioVision' meeting in Lyon, France, the G8 leaders promised in 2005 to double aid to Africa from US$25 billion in 2004 to at least $50 billion in 2010. But African countries still have no idea how or when aid levels will increase. Creative accounting means that supposed new aid is sometimes just a repackaging of existing aid or a debt cancellation. What aid has emerged has not led to the organized, massive expansion of investments in clinics, schools, agriculture and infrastructure that is needed.

This bad situation is made even worse by the pseudoscientific veneer conferred by evaluating progress on the MDGs using 48 quantitative indicators compared with a 1990 baseline. Every year, the UN rolls out reports with slick graphics, seemingly noting with precise scientific precision progress towards the goals. But the reports mask the fact that the quality of most of the underlying data sets is far from adequate. Moreover, the indicators often combine very different types of data, making aggregation and analysis of the deficient data even more complicated.

There are decent data for just a handful of indicators, such as child mortality, but for most of the 163 developing countries, many indicators do not even have two data points for the period 1990–2006. And few developing countries have any data for around 1990, the baseline year. It is impossible to estimate progress for most of the indicators over less than five years, and sparse poverty data can only be reliably compared over decades. To pretend that progress towards the 2015 goals can be accurately and continually measured is false.

Significant efforts are now being made to improve data collection. Meanwhile, UN agencies fill in the missing data points using 'modelling' — in practice, a recipe for potentially misleading extrapolation and political tampering.

Indeed, the lack of data makes it impossible not only to track progress, but also to assess the effectiveness of measures taken. Has the existence of the MDGs changed pre-existing trends? Are bednets helping to control malaria? Are improvements in Asia down to the MDGs or simply economic growth? Currently, it's impossible to tell. Meanwhile, spurious claims of achievement are promoted.

Funding the scientific evaluation of interventions would pay dividends in enabling rigorous project management. But although billions of dollars are now flowing into aid and disease control, researchers complain that they struggle to get even tiny funds for evidence-based research to assess which interventions work. "If I want 10 tonnes of DDT it's no problem; if I want $10,000 to see if the 10 tonnes made any difference, forget it," says one malaria researcher.

"A lack of data makes it impossible not only to track progress, but also to assess the effectiveness of measures taken."

It is important to take action towards the goals rather than use the lack of reliable information as an excuse for inaction. But investment in an evidence-based approach to aid interventions, assessed independently of the UN, is also essential. Otherwise, in 2015, the MDGs could be buried in history's graveyard alongside other well-intentioned but failed development efforts. ■

Independence day?

Spain has increased science funding but now needs to modernize the organizations at the top.

'New government pledges to double research budget.' How often have scientists heard this type of promise and been disappointed? All the more credit, then, to Spain's politicians, who have delivered on it. When the socialist government swept to power in March 2004, it launched a multifaceted programme called Ingenio 2010, founded on a commitment to raise the country's spending on research from 1.1% of its GDP — well below the European Union average of 1.8% — to 2% by 2010. The civilian research budget has already more than doubled.

Given equivalent budgets, would Spain's researchers be able to deliver the same quality as the big European players? Some of Spain's élite research institutes, such as the CNIO, the national cancer centre in Madrid, already do so. But broader success will depend on Spain's plans to modernize its science management. If the money is to be well spent, the new basic-research agencies currently being created must adopt European Union norms. In particular, they must be free of direct political interference — a point of contention among some government bureaucrats.

Researchers say the new money is already making a palpable difference. The number of research positions, the acceptance rate of grants, and grant sizes have all increased. There are more diverse sources of grants, thanks to strategic programmes opened up within

Ingenio 2010. There is significantly more money for infrastructure, and astronomers and particle physicists will gain from Spain's recent membership of the European Southern Observatory and higher subscription to CERN. The budget of the CSIC, the national research organization, which runs 115 research institutes and centres, has seen a healthy 74% increase.

The country seems well placed to absorb the new money. Despite very modest investment over the previous two decades, the quantity and quality of Spanish publications have increased markedly. The bad news is that Spain's science organization is stuck in a time warp. Both the government's research funding department and the CSIC are embedded in a slow and bureaucratic research ministry, where, it is said, a grant application arriving in the first-floor mailroom can take weeks just to arrive in the correct department a few floors higher. Scientists paid from the public purse — at the CSIC or at universities — are civil servants, hired for life and virtually impossible to fire. The positions, like the detailed budgets, are fixed centrally.

Recognizing the need for more flexibility, the government last year passed a law allowing certain ministerial activities to be spun out into more independent agencies. Two of these, now in a rather painful process of creation, will be the CSIC and an agency for research funding, evaluation and foresight.

The new CSIC concept, steered by its president, immunologist Carlos Martínez, will seek government approval in the coming weeks. It allows the CSIC to be run by its president, who must be a respected scientist, with arms-length oversight by government representatives, and with the opportunity to offer scientists negotiable contracts. This change will enable the CSIC's institutes to compete internationally.

This proposal has to make its way through other government departments, no doubt nervous of delegating responsibility for large amounts of public money. It is nevertheless imperative that it emerges as an appropriate model for others to follow. It is equally important that the concept for the research funding, evaluation and foresight agency is accelerated so it can be put in place before Spain's general election in March next year.

Until Spain has these independent and flexible research structures safely in place, the success of the well-conceived Ingenio 2010 programme is threatened. Failure would be a heavy price to pay for a lack of imagination, especially when the required independence plays such a key part in success elsewhere. ■

Going underground

A new study's recommendations for carbon capture should be pursued.

The world has no shortage of coal: about a trillion tonnes of the stuff is considered to be recoverable worldwide. It's also cheap: joules from coal cost considerably less than those from natural gas. But coal has a deservedly bad reputation for carbon emissions, which are greater per kilowatt-hour than for any other major electricity source.

The United States has a quarter of the world's coal, and half of its electricity is coal-generated. But power companies are finding it increasingly difficult to build new coal plants, even with increasing demand. The difficulty is twofold. Environmental lawyers and lobbyists are making the process as lengthy, expensive and unpredictable as possible. And companies are increasingly convinced that carbon regulation is coming to the country, maybe not today, maybe not tomorrow, but soon, and for the rest of our lives. Politicians in the Democratic Party have made clear their intention of penalizing any attempt to cheat future regulations by building large numbers of plants before these regulations take effect.

This confluence of factors contributed to the decision this year by Texas electricity company TXU to back away from its idea to build eleven new coal plants (see page 362). Instead, the company looks set to be bought out by a group of investors who cannily obtained public support for the deal from environmental groups, in part by pledging not to build eight of the eleven plants — a decision they would probably have made anyway. Meanwhile, the idea of banning further construction of coal plants is being promoted by an unprecedented range of people, including NASA's Jim Hansen, members of the Texas legislature, and a United Nations panel chaired by the director of the Missouri Botanical Garden, Peter Raven. The epoch of US coal plants without carbon capture and storage may be coming to an end.

That would be just fine for many people, to whom coal is anathema. There are plenty of alternatives, from renewables to natural gas. The latter is plentiful in the United States and is cleaner-burning. And although the general idea behind carbon capture and sequestration seems reasonable to most experts, it hasn't been tried at the truly colossal scales that would be needed. Pumping gigatonnes of carbon underground is not to be undertaken lightly. Early studies indicate that there is sufficient capacity, but detailed surveys must be made. Questions about leakage and the best ways to monitor the sites must be answered. Full-scale demonstration plants must be built with varying technologies and on different geologies, and they must be carefully monitored and studied.

This is challenging indeed, but a study by researchers at the Massachusetts Institute of Technology, *The Future of Coal*, finds that the technology is viable. According to their models, it will be adopted when carbon emissions cost US$30 a tonne. The report calls for immediate investment in carbon capture and sequestration projects, with a total of ten projects worldwide at a cost of about $15 million each per year for ten years.

"The MIT report calls for immediate investment in carbon capture and sequestration projects."

This figure of $1.5 billion may sound big, but it's tiny when compared to the role of coal in the worldwide energy industry. Despite all the uncertainties and all the other energy options, this proposal should be implemented. If carbon markets and competing technologies eventually lead most countries to abandon coal as an energy source, so much the better. But it is impossible to be sure of such an outcome, and the risks of global change are serious enough for the United States and other countries to be pursuing all low-carbon energy sources with vigour. ■

RESEARCH HIGHLIGHTS

Feel the heat

PLoS ONE doi:10.1371/journal.pone.0000281 (2007)

Crocodiles rely on temperature-sensing ion channels in their cells to decide whether to warm up in the sun or take a cooling swim, say Frank Seebacher and Shauna Murray of the University of Sydney, Australia.

The researchers found that *Crocodylus porosus* (pictured) has two genes, *TRPV1* and *TRPM8*, that are closely related to genes found in warm-blooded animals. The genes encode ion-channel proteins that act as heat and cold sensors. Inactivating the sensors impairs the reptile's ability to thermoregulate: crocodiles whose sensors were chemically blocked stopped shuttling between sunbathing and swimming, and their body temperatures settled down to that of the water.

D. ZUPANC/NHPA

CELL BIOLOGY

Make or break

Cell **128,** 915–929 (2007)

Researchers have uncovered a molecular link between the construction and clear-up of scaffolding inside a migrating cell.

The skeleton of a migrating cell must constantly re-form, with actin filaments being assembled at the leading edge and broken down at the rear. James Bear of the University of North Carolina, Chapel Hill, and his colleagues suggest that a protein called coronin 1B coordinates these processes.

They found that coronin 1B inhibits actin-filament formation by interacting with a protein complex known as Arp2/3, and that it also regulates the activity of the protein cofilin, required for filament turnover. Furthermore, depleting coronin 1B in rat cells reduced their motility.

MATERIALS SCIENCE

Love–hate relationship

Macromolecules doi:10.1021/ma062965u (2007)

Here's a riddle: how can a 'hydrogel' that holds more than its dry weight in water have a hydrophobic, or water-repellent, surface?

Kazutoshi Haraguchi of the Kawamura Institute of Chemical Research in Sakura, Japan, and his colleagues observed this for a hydrogel made from a network of the polymer poly(*N*-isopropylacrylamide) and particles of clay.

Water dropped onto a hydrogel would usually spread out and be absorbed. But water splashed onto the surface of this team's material sat in round droplets — making contact angles of up to 150°. The researchers suspect that the polymer's water-repellent side-chains have wriggled to the surface, poking out to form an impenetrable layer.

CHEMISTRY

The buckycatcher

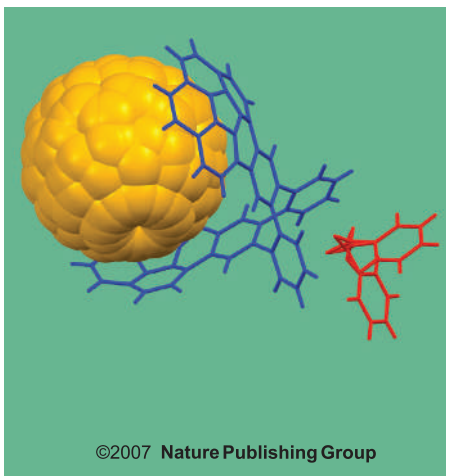
J. Am. Chem. Soc. doi:10.1021/ja070616p (2007)

Molecular tongs designed to clasp carbon C_{60} molecules — or buckyballs — make their debut in a paper from Andrzej Sygula of Mississippi State University and his team.

The tongs have two gently curved, bowl-shaped heads that wrap snugly around C_{60}'s spherical shell, cradling it in the same way a baseball glove holds a ball (pictured below).

The bowls are molecules known as corannulenes, which contain aromatic rings of carbon atoms. They bind C_{60} through interactions between the two structures' π molecular orbitals.

Attached to a surface, these tongs could be just the thing for extracting C_{60} molecules from solution, or for positioning them at precise locations.



DEVELOPMENTAL BIOLOGY

Secure checkpoint

Curr. Biol. doi:10.1016/j.cub.2007.02.027 (2007)

Fruitflies use a belt-and-braces approach to stop genetic parasites wrecking the chromosomes in their developing eggs, report Trudi Schüpbach and her team at Princeton University, New Jersey.

The team found that mutations in genes involved in the 'rasiRNA' pathway — a kind of RNA interference — led to a marked increase in the activity of mobile DNA elements called transposons. This suggests that the rasiRNA pathway normally shuts down transposons, whose activity destabilizes chromosomes.

The transposon activity then prompted a protein known as Chk2, which responds to DNA damage, to either halt cell division in the developing eggs, or, if that failed, block the translation of a protein essential for egg development. This 'checkpoint' monitors the RNA-interference system and kicks in if it fails, the team suggests.

INORGANIC CHEMISTRY

Radical element

Angew. Chem. Int. Edn doi:10.1002/anie.200700059 (2007)

Chemists have achieved a delicate balancing act: creating a molecular 'radical' containing a phosphorus atom with an unpaired electron that will react with other species, but which is stable enough to be stored in a bottle. Such a radical might be used to initiate polymerization reactions, or to probe reaction intermediates.

Phosphorus radicals are typically difficult to stabilize. Here, the phosphorus

is sandwiched at the radical's core between two chemical groups that contain vanadium atoms. This stabilizes the structure, report Christopher Cummins at the Massachusetts Institute of Technology in Cambridge and his team, because the phosphorus shares its lone electron with the two metal atoms. They hope that the same strategy will stabilize radicals centred on other elements, and that varying the metal will tune the radical's reactivity.

MALARIA

Which mozzies win out?

Proc. Natl Acad. Sci. USA **104,** 5580–5583 (2007)

Genetic resistance to the malaria parasite gives mosquitoes feeding on infected blood a fitness advantage, researchers have found. Release of mosquitoes that are resistant to infection with malaria is one control strategy being considered to curb the disease.

Marcelo Jacobs-Lorena of the Johns Hopkins University in Baltimore, Maryland, and his colleagues put 250 transgenic and 250 wild-type mosquitoes of opposite sexes into a cage, where they fed on mice infected with the *Plasmodium berghei* parasite. The transgene, which blocks infection through the mosquitoes' gut, was found in around 70% of the mosquito population after 10 or so breeding cycles. Mosquitoes in the wild only occasionally become infected with the parasite, but this study gives hope that the transgene could persist in the population.

ASTRONOMY

Seeing things

Astrophys. J. **657,** 669–680 (2007)

Infrared light thought to have been emitted by the Universe's first stars isn't seen in a new survey of the skies.

Researchers had previously found a 'near-infrared background excess' in some

NASA/ESA/R. THOMPSON

satellite images that they couldn't account for with known sources. They argued that it was light from early galaxies, stretched by the expansion of the Universe to appear at infrared wavelengths.

Rodger Thompson of the University of Arizona, Tucson, and his colleagues analysed sharper and more sensitive images from the Hubble telescope (pictured above). They say the claimed excess was due to inaccurate estimates of emission from zodiacal dust. What's more, they could attribute spatial variations in the background to previously undetected nearby galaxies.

NANOTECHNOLOGY

Spheres inside cells

Environ. Sci. Technol. doi:10.1021/es062541f (2007)

Concerns about nanoparticle toxicity have prompted researchers to look closely at how C_{60} molecules interact with cells.

Alexandra Porter at the University of Cambridge, UK, and her colleagues imaged C_{60} that had infiltrated human macrophages — cells that have a role in clearing debris from the lungs. The researchers showed that a technique known as energy-filtered transmission electron microscopy can pick out the carbon spheres. They could see individual molecules and tell apart aggregates that were crystalline or disordered.

C_{60} appeared in the cells' cytoplasm and nuclei. The molecules were concentrated just inside the cell wall, suggesting that they had infiltrated the cell through its membrane.

GENETICS

Mutations linked to autism

Nature Genet. **39,** 319–328 (2007)
Science doi:10.1126/science.1138659 (2007)

Two studies point to genetic changes that may contribute to the spectrum of autism disorders.

The Autism Genome Project Consortium scanned the genomes of more than 1,000 affected families for single-nucleotide variations in DNA inherited alongside the disorder. They also searched for inherited versions of mutations known as copy-number variants — deletions or duplications of chunks of the genome. Their findings implicate two major regions of DNA, one of which is linked to neuronal proteins called neurexins.

Independently, Jonathan Sebat and Michael Wigler at Cold Spring Harbor Laboratory, New York, and their colleagues compared the role of copy-number variants in sporadic and inherited cases of autism. They found that such mutations appear spontaneously in 10% of patients with sporadic autism, but in only 2% of patients from families with more than one affected member. This suggests that the two classes of autism differ in the primary genetic mechanism involved.

Correction

The Research Highlight 'An added dimension' (*Nature* **446,** 234; 2007) incorrectly referred to *Schizosaccharomyces pombe* as budding yeast. It is fission yeast.

JOURNAL CLUB

James Bauer
College of William and Mary, Gloucester Point, Virginia, USA

A marine scientist marvels at connections between the cold war and slimy mudflat worms.

Having grown up on the coast of New England, my childhood involved a good deal of digging around in the intertidal mud, unearthing things that most people of good sense do their best to avoid — things such as slimy, slithering worms, which often bite or smell bad, or both.

Older but no wiser, I was delighted to come across a recent paper (E. Teuten *et al. Mar. Ecol. Prog. Ser.* **324,** 167–172; 2006) that has cleverly extracted a surprising scientific result from studies of such mudflat worms.

As well as reminding me of my dubious childhood pastime, the work recalls the period in which I grew up, during the cold war, when much of the world lived in fear of the nuclear weapons then being tested. This work takes advantage of one legacy of those tests.

The bomb tests sent into the atmosphere lots of the isotope carbon-14, normally present only at low levels. This bomb carbon-14 subsequently made its way into the oceans, where it became incorporated into plankton. The plankton in turn sank and became part of the coastal mud, providing a home and a food source for marine sedimentary animals.

Mudflat worms are generally believed to ingest wholesale the nondescript sediment in which they live, yet the worms examined in this study contained more bomb carbon-14 than the sediment surrounding them.

Thus, it seems that the worms assimilate from the amorphous goop, material that has been deposited since the cold war and so is younger than the average age of the sediment. Presumably, they do so because the newer material is more nutritious, but how they extract it is unknown.

Makes me want to get back out by the sea with my bucket.

SPECIAL REPORT

Degrees in homeopathy slated as unscientific

Alternative therapies are now a degree subject at some British universities. But do they deserve these credentials? **Jim Giles** reports.

As debate rages in the United States over whether intelligent design should be taught in science classes, another topic that many researchers see as a pseudoscience is claiming scientific status within the British education system.

Over the past decade, several British universities have started offering bachelor of science (BSc) degrees in alternative medicine, including six that offer BSc degrees in homeopathy, a therapy in which the active ingredient is diluted so much that the dose given to the patient often does not contain even a single molecule of it. Some scientists are increasingly concerned that such courses give homeopathy and homeopaths undeserved scientific credibility, and they are campaigning to get the label removed (see Commentary, page 373).

Many scientists and advocates of evidence-based medicine feel that giving homeopathy scientific status is unjustified. Aside from the fact that there is no known mechanism by which this treatment could work, they argue that the evidence against it is conclusive. Of the many rigorous systematic reviews conducted in the past decade, only a handful have produced evidence, marginal at best, in favour of homeopathy, with the authors in each case stating that the data were weak. Several reviewers found no effect, and a prominent study suggesting that homeopathy does work (L. Linde *et al. Lancet* **350,** 834–843; 1997), and which is frequently cited by homeopaths, has had its methodology extensively criticized since publication.

"I can only imagine they teach it's OK to cherry-pick evidence. That's totally unacceptable."

But homeopaths involved in the university courses — those that were willing to speak to *Nature*, at least — argue that they teach students scientific principles, including the critical analysis of evidence.

Finding out exactly what is taught in the courses is not straightforward. Ben Goldacre, a London-based medical doctor, journalist and frequent critic of homeopathy, says that several universities have refused to let him see their course materials. "I can't imagine what they're teaching," he says. "I can only imagine that they teach that it's OK to cherry-pick evidence. That's totally unacceptable."

Pharmacologist David Colquhoun of University College London has had the same problem, and is now using freedom-of-information legislation to get access to course materials after having numerous requests refused. The University of Central Lancashire and the University of Salford both declined requests to talk to *Nature* or share details of their homeopathy degrees.

One university that is willing to discuss its teaching is the University of Westminster in London. Brian Isbell, head of Westminster's department of complementary therapies, defends the BSc description, arguing that as with all of the university's complementary therapy degrees, students also have to study the health-sciences model of disease, so that they can "work safely and effectively within the healthcare system". Students are required to do research and produce critiques of the literature. Reading lists include papers from sources such as *Homeopathy*, a journal published by the Faculty of Homeopathy, a members' association for professional homeopaths based in Luton. But the lists also include recent studies that are critical of homeopathy and conventional guides to doing and evaluating healthcare research.

One assignment asks students to critique a paper that assessed the health changes reported by patients suffering from a range of chronic diseases when they attended follow-up appointments after receiving homeopathic treatment at a hospital in Bristol. Almost three-quarters of the 6,500 patients reported that their condition had improved (D. S. Spence *et al. The Journal of Alternative and Complementary Medicine* **11,** 793–798; 2005).

The paper generated significant media coverage when it was published, but its methodology has been widely criticized. No control group was used, prompting Colquhoun to note that the study is not even capable of showing that homeopathy was producing a placebo effect. So what happens when students critique the paper? Do they get full marks for showing that it provides no evidence at all for homeopathy?

Not quite, says Isbell, who says the paper was chosen precisely because of the controversy over its methodology. Students would be expected to discuss the problems with the lack of controls and to suggest ways to run better studies. But Isbell says that the Bristol researchers still collected useful 'outcome measures' — basically a set of reports from individual patients about how they improved. "It doesn't have the rigour of other methods," Isbell says, "but it is part of the picture."

Homeopathic medicine is big business, but giving it the status of a science degree is controversial.

The differing opinions over the paper highlight an issue at the centre of the dispute about the evidence for homeopathy, and which explains in part why lecturers feel they can teach the subject as science. For advocates of evidence-based medicine, the double-blind randomized clinical trial (in which neither the doctor nor the patient knows who is getting active treatment and who is getting a placebo) is the best form of evidence available to practitioners. When regulators take decisions on drug safety, for example, they usually rely on such studies. But for homeopaths, there is a serious flaw in this approach.

When a patient visits a homeopath, the practitioner asks questions that go beyond the symptoms and probe other aspects of the

P. MACDIARMID/GETTY

patient's life, such as whether they are feeling stressed or unhappy. The result is an individualized treatment that takes longer than the ten or so minutes that the patient would get with a government-funded family doctor. This personal interaction is critical to homeopathy, both in tailoring the medicine and in gaining the patient's confidence. Homeopaths say that if there is a chance that the patient might receive a placebo at the end of it, the necessary trust can break down.

"Trying to do what I do in that context didn't work very well," says Clare Relton, a practising homeopath who is conducting research into homeopathy at the University of Sheffield and has taken part in a clinical trial designed to assess homeopathic treatments for chronic fatigue syndrome. "I found it difficult to build a therapeutic relationship," she says. Relton argues that homeopathy is scientific, but that the problem of trust means that double-blind trials aren't the best way to measure its effectiveness. Instead, she and other homeopaths prefer to rely on more qualitative methods, such as case studies and non-blinded comparisons of treatment options. These, they say, provide ample evidence that homeopathy works.

Similar attitudes to homeopathic teaching are evident outside Britain. "I definitely think homeopathy can, and should, be evaluated within the context of double-blind, placebo-controlled, randomized clinical trials," says Ellen Hughes, who teaches complementary therapies to medical students at the University of California, San Francisco. She concedes, however, that the emphasis homeopaths place on individually tailored treatment makes designing such trials "a bit of a challenge".

For advocates of evidence-based medicine, such arguments are equivalent to admitting that homeopathy is nothing more than a strong placebo effect brought on by an attentive practitioner. If the treatment cannot work unless the patient and practitioner believe in it, then it cannot be due to the physical properties of the remedy. Homeopaths disagree, insisting that the remedy itself does have an effect independent of the practitioner. But by ruling out what scientists consider the best mechanism available to test this assertion, it is hard to see how homeopaths will ever convince their opponents.

So where does this leave scientific opposition to homeopathy degrees? Outside Britain, even in France and Germany where homeopathy is relatively popular, the topic is taught in universities only as a small part of a medical degree and is not classed as a science — a compromise that most academics seem happy with.

But in Britain, the number of BSc degrees in alternative medicine has grown over the past decade. They are generally run by 'new' universities — institutions that emphasize vocational rather than academic training, but have been given university status over the past 15 years as part of the government's drive to provide equal opportunities for higher education. Alternative medicine is not the only surprising subject to be classified as science (see "A science subject?"), but Colquhoun and Goldacre argue that degrees in complementary medicine are particularly harmful because they lead patients to believe that they are being treated by a scientifically trained practitioner.

The critics seem to have little chance of getting the BSc label removed from these courses any time soon. The few organizations that could pressure universities to reclassify the courses have little interest in the debate. Universities UK, the body that represents the country's higher-education institutes, says that it has not discussed the matter and that decisions about how to describe courses are up to the individual universities. The Quality Assurance Agency for Higher Education, the body charged with safeguarding academic standards, also says that it does not get involved in questions about what constitutes science, and that universities are entitled to set their own courses.

So *Nature* contacted the universities of Westminster, Central Lancashire and Salford for an official response from the institution on whether they think the BSc tag is justified for their homeopathy courses. All declined to comment. ■

Additional reporting by Declan Butler, Michael Hopkin, Katharine Sanderson and Sophie Sigler.

A science subject?

You can earn a BSc degree in quite a few subjects that wouldn't conventionally be called science. Here are a few:

Geography with Mountain Leadership
Staffordshire University

Hospitality Management
Manchester Metropolitan University

Police Studies
Buckinghamshire Chilterns University College

Adventure Recreation
Harper Adams University College, Newport

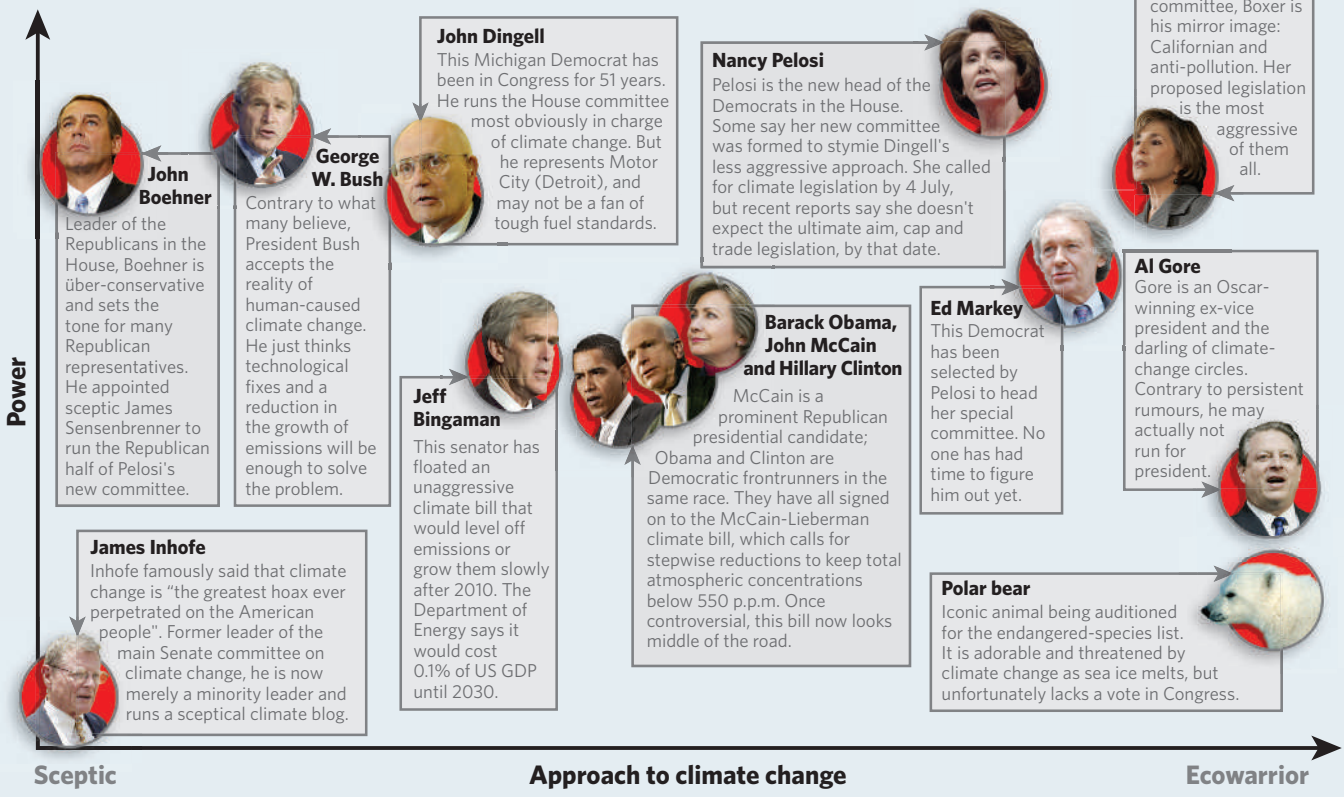
Community Health and Leadership Studies
Sheffield Hallam University

Sports Event Management
Leeds Metropolitan University

GRAPHIC DETAIL

Where politicians stand on climate change

Since the Democrats took over the US House and Senate, ever more committees on climate change are holding hearings on a nearly daily basis. This month the leader of the House, Nancy Pelosi, formed yet another one, the Select Committee on Energy Independence and Global Warming, leaving even Washington energy analysts struggling to keep up with it all. **Emma Marris** charts the key players in terms of clout and greenness.



D. COOK/AP; J. ERNST/REUTERS; R. L. WOLLENBERG/UPI/NEWSCOM
S. J. FERRELL/CONGRESSIONAL QUARTERLY/NEWSCOM; M. SEGAR/REUTERS; S. WALSH/AP; D. MCNEW/GETTY
M. JALIL/AP
S. SCHLIEBE/USFWS; R. SULLIVAN/AFP/GETTY; S. SCOTT/UPI/NEWSCOM; P. SAKUMA/AP; K. DIETSCH/UPI/NEWSCOM

Agencies join forces to share data

The US government is considering a massive plan to store almost all scientific data generated by federal agencies in publicly accessible digital repositories. The aim is for the kind of data access and sharing currently enjoyed by genome researchers via GenBank, or astronomers via the National Virtual Observatory, but for the whole of US science.

Scientists would then be able to access data from any federal agency and integrate it into their studies. For example, a researcher browsing an online journal article on the spread of a disease could not only pull up the underlying data, but mesh them with information from databases on agricultural land use, weather and genetic sequences.

Nature has learned that a draft strategic plan will be drawn up by next autumn by a new Interagency Working Group on Digital Data (IWGDD). It represents 22 agencies, including the National Science Foundation (NSF), NASA, the Departments of Energy, Agriculture, and Health and Human Services, and other government branches including the Office of Science and Technology Policy.

The group's first step is to set up a robust public infrastructure so all researchers have a permanent home for their data. One option is to create a national network of online data repositories, funded by the government and staffed by dedicated computing and archiving professionals. It would extend to all communities a model similar to the Arabidopsis Information Resource, in which 20 staff serve 13,000 registered users and 5,000 labs.

The group then aims to help scientific communities create standards to let databases in one field talk to others in different disciplines. That will be no mean feat. Even scientists in highly organized 'big science' fields such as genomics and astronomy, who routinely work with large shared data sets, encounter problems when trying to carry out sophisticated calculations using data from different communities.

For example, it is still a huge task to mash together astronomical observations of the same object taken at different wavelengths and stored in different databases, says Giuseppina Fabbiano of the Harvard-Smithsonian Center for Astrophysics, the Smithsonian's representative on the IWGDD. "We still cannot extract data from different archives and put it together seamlessly." Agreeing standards even within closely linked communities "is like Middle-East peacekeeping — every detail has to be worked out and agreed on," adds her colleague at the centre, Martin Elvis.

"Agreeing standards is like Middle East peacekeeping — every detail has to be agreed on."

Many researchers are reluctant to share their raw data in the first place. The IWGDD is considering making submission of well-documented data sets to archives a requirement of getting a grant.

Christopher Greer, senior adviser for digital data at the NSF's Office of Cyberinfrastructure, says that if and when all federally supported science data are accessible, he hopes that publishers and computing companies will add on more sophisticated information services. This would give researchers unprecedented ability to test their ideas: "The next web browser could be a visualization and data-navigation tool, and the next Google an information integrator." ■

Declan Butler

Looking for hidden signs of consciousness

Can brain-injured patients who show no response to their surroundings ever be considered conscious? This question became a hot topic last year after researchers who scanned the brain of a woman diagnosed as being in a vegetative state found that she could perform certain mental tasks on request.

The same team has now validated its scanning method on healthy volunteers, and believe that it can be used as a general method for measuring consciousness in unresponsive patients. The researchers are also using real-time brain scanning to try to ask questions of patients who pass the test.

Given that there is no consensus about what it even means to be conscious, it is perhaps not surprising that it is difficult to find an objective way of detecting any trace of it in patients in a vegetative state.

A test is needed, nevertheless, says Steven Laureys, a team member at the University of Liège in Belgium. Some signs of consciousness have been found in up to 40% of patients in an apparent vegetative state when they are studied more closely, suggesting that such patients are often misdiagnosed.

In the high-profile case reported in *Science* last year, the team, led by Adrian Owen of the MRC Cognition and Brain Sciences Unit in Cambridge, UK, used functional magnetic resonance imaging (fMRI) to show that a woman left in a vegetative state after a car accident could respond to requests to imagine playing tennis or navigate around her house (A. Owen *et al. Science* **313,** 1402; 2006).

The team has now tested this technique on 24 healthy volunteers, who were similarly instructed to imagine either walking around their house or playing tennis. The tasks activate separate networks in the brain, and the scans proved able to tell correctly which task was being performed (M. Boly *et al. NeuroImage* doi:10.1016/j.neuroimage.2007.02.047; 2007).

The researchers say that showing that the method works reliably in healthy brains proves its robustness. "Our challenge is to find markers that tell us 'this is a hopeless case' or 'this is a case where we should increase our therapeutic efforts,'" says Laureys.

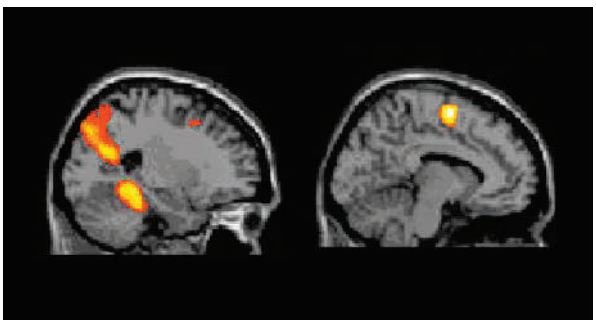
> "Our challenge is to find markers that tell us whether we should increase our therapeutic efforts."

Laureys has since used the technique in five vegetative-state patients in Liège, but none has shown brain activity compatible with consciousness. Owen has tried with three patients in Cambridge — and found one man capable of responding. His team went on to try two-way communication with this patient and with the girl originally described in *Science*, using real-time brain scanning. The patients were instructed that to answer 'yes' to a question they should imagine playing tennis (or football — the team knew the man was an avid Liverpool fan), and for 'no' to imagine walking around their house.

This is considerably more complex than the original task and neither patient has responded, but Owen is confident that some patients will be capable of doing so. "It's probably only a matter of time," he says. Because doing fMRI scanning requires transporting patients to a hospital that has a scanner, Owen is also investigating the idea of fitting patients with electroencephalogram (EEG) caps that would measure their brainwaves and use this activity to move a cursor around a computer screen. Patients who passed the initial fMRI test could then be extensively followed up using EEG.

Not everyone is convinced that the test is ready for diagnostic use. Even if patients do not respond, cautions Nicholas Schiff, a neurologist at Weill Cornell Medical College in New York, it is still not possible to tell exactly what this means. Does the patient understand the question? Do they not want to respond? "This is not ready for primetime," says Schiff. ■

Kerri Smith



NEUROIMAGE, ELSEVIER

Imagining spatial navigation (left) and playing tennis.

ON THE RECORD

"You're a big daft cock."

The entire response of Martin Durkin, director of a recent British television documentary *The Great Global Warming Swindle*, to a reasoned e-mail from Imperial College scientist Armand Leroi questioning the programme's accuracy.

SCORECARD

Being fat
Obese men are less prone to suicide, according to a 20-year survey, although the study's authors do not recommend overeating as a healthy route to happiness.

Living on the Moon
NASA has set up a working group to investigate whether toxic dust could make 'lunar hay fever' an occupational hazard of living on a Moon base.

ZOO NEWS

FOTOLINCSV/ALAMY

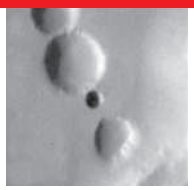
Pigeon-fanciers in Western Europe are looking forward to 1 April, when a ban imposed to reduce the risk of bird flu is set to be lifted.

Wildlife managers in Western Australia are drawing up plans to control feral camels driven "mad with thirst" by the recent drought, causing them to rampage through towns and farms in search of water.

Sources: ocean.mit.edu, Arch. Int. Med., BBC, Environment News Service, Desert Knowledge Cooperative Research Centre

A. & S. TOON/NHPA

SIDELINES



CAVES SPOTTED ON MARS
Heat signal identifies caverns on the red planet.
www.nature.com/news

A jump that would prove Newton wrong

On 21 March — the spring equinox — something very strange may have happened. In two particular places on Earth, objects might have started to move without any force acting on them.

Such motion would violate Newton's second law, a fundamental principle of mechanics, which states that objects accelerate by an amount proportional to the force acting on them. But it needn't involve anything paranormal. According to Alexander Ignatiev of the University of Melbourne in Australia, this weird phenomenon would be proof of a new kind of mechanics called modified newtonian dynamics (MOND).

First suggested in 1983 by Israeli physicist Moti Milgrom, MOND aims to provide an alternative to dark matter, the invisible and so far unidentified stuff invoked by physicists to explain why rotating galaxies don't fall apart.

MOND is very speculative, but interest in it has been boosted in recent years by the observation that the Pioneer 10 and 11 spacecraft, launched in the early 1970s to study the planets, seem to be veering from their expected paths as they leave the Solar System. No one knows whether the anomaly is real or just the result of faulty observations of the spacecraft's motion.

Finding evidence for MOND would be hugely important: "The foundations of physics would have to be revised," says Ignatiev. But measuring the predicted deviations from Newton's theory is extremely difficult. MOND diverges from newtonian dynamics only at very small accelerations, around a hundred trillionths of a metre per second per second. At this rate, an object that began to move when Newton published his second law in 1687 would now be travelling at one metre per second.

To check out the theory, Ignatiev says, you need a test object moving with an acceleration small enough (relative to the rest of the Galaxy) for MOND effects to be apparent. Even if an object sits motionless on a lab bench, it is generally accelerating with respect to the Galaxy because of Earth's rotation and orbit around the Sun, and the Sun's motion in space.

But Ignatiev's calculations have shown that at a particular moment each year, at two points on Earth, these accelerations cancel out for about half a millisecond. This, he calculates, coincides with either the spring or autumn equinox (A. Yu. Ignatiev *Phys. Rev. Lett.* **98,** 101101; 2007). The points of cancellation are always at latitudes 79° 50′ above and below the Equator, whereas the longitude varies each year. This spring they will be at 178° E, high above Siberia in the Arctic Sea and on the Ross Ice Shelf of Antarctica. The autumn equinox of 2008 will be a little more amenable to experimenters, when the northern point is in Greenland.

A spooky event violating Newton's second law might have occurred on Antarctica's Ross Ice Shelf.

R. PRICE/PHOTOLIBRARY.COM

Ignatiev cautions, however, that these are approximate locations, and ignore the subtle effects of the Moon and planets. Before you make the journey, he says, you had better do the full calculation, as the effect acts over an area just a few centimetres across.

So what might you hope to see? As external acceleration effects cancel out, says Ignatiev, the object's acceleration would fall below the MOND threshold — and if the theory is correct, it will spontaneously jump over a small distance.

> **"The foundations of physics would have to be revised."**

The predicted jump is tiny: about 0.2×10^{-16} metres, or one-fiftieth of the diameter of a proton. But Ignatiev reckons it could be measured by interferometry, which detects minute differences in the path lengths of light beams. The technique is being developed in huge instruments for detecting gravitational waves, which are predicted to alter the dimensions of space very slightly as they pass. Gravitational-wave detectors typically involve light corridors several kilometres long. But there are now other types of detector that are more portable, and which it might be possible to carry to Antarctica or Greenland. The MiniGRAIL detector being developed at Leiden University in the Netherlands, for example, is a metal ball 68 centimetres in diameter.

Others in the field are sceptical about Ignatiev's chances. Orfeu Bertolami, a physicist at the Instituto Superior Técnico in Lisbon, says that although the proposal is "a brave attempt" to suggest how MOND could be tested, he is not convinced it would work. He points out that Ignatiev used newtonian mechanics to calculate the points of zero acceleration, and that these rules wouldn't necessarily apply in the MOND world he wants to test.

Even if the theory works, Bertolami doubts it would be possible to observe exact cancellation of acceleration. A medium-sized iceberg passing 10 kilometres away in the Antarctic or Arctic Ocean, he says, would induce a gravitational acceleration comparable to the MOND threshold.

And the MOND hypothesis itself is looking decidedly shaky, because there is now rather good evidence that dark matter, which MOND attempts to eliminate, does exist. Last year, a galaxy cluster was found to have an asymmetrical distribution of visible matter that could only easily be explained if it was balanced by dark matter (see *Nature* doi:10.1038/news060821-6; 2006). After this, says Bertolami, "any attempt to force MOND on us is difficult to swallow".

But Ignatiev argues that the debate is far from settled. "Experimental searches for dark matter are conducted in many laboratories across the world," he says. "MOND should be given the same chance." ■

Philip Ball

Tamiflu side effects come under scrutiny

DAJ/GETTY

TOKYO

Japanese paediatricians are studying whether Tamiflu, the drug widely thought to be our best defence against avian flu, might be causing mental instability and suicidal tendencies. An initial study cleared Tamiflu of any link, but that study, as well as a larger one that has just started, are raising concerns about conflicts of interest, as some of the researchers have received large sums of money from the drug's Japanese distributor.

Countries around the world have been stockpiling Tamiflu (oseltamivir) since 2004, in case of a pandemic of avian flu. Japan is by far the biggest user of the drug: in 2005, Japanese doctors wrote 9 million prescriptions for Tamiflu, compared with 3 million for all other countries combined. But concerns about the drug's safety, which is made by Roche, have been growing after some unusual deaths in Japanese teenagers who had taken it. Most recently, a 14-year-old girl and a 14-year-old boy died when they jumped from apartment buildings on 16 and 27 February, respectively.

Tokyo-based Chugai Pharmaceuticals, which distributes the drug in Japan, says that it has reported 289 cases of psychoneurotic effects, including three suspicious deaths, to the Japanese health ministry since the drug was launched there in 2001. Chugai now lists the possibility of severe neurological side effects on the drug's labelling and has distributed a warning note to hospitals. The Japanese health ministry insists that there is no clear evidence of a link, however, and points out that flu itself can cause symptoms such as abnormal behaviour. Roche agrees, adding that the number of suspicious deaths is tiny compared to the number of people who have been prescribed the drug.

Tamiflu was cleared initially when seven paediatricians and a statistician looked at the side effects of it and other influenza drugs, as well as at the symptoms of flu itself. The study, which ended in February 2006, followed up 2,846 children aged mostly ten or younger who had been diagnosed with flu. The frequency of abnormal behaviour in children who took Tamiflu was 11.9%, compared with 10.6% in those who didn't take it, which the researchers concluded was not a significant difference.

Shunpei Yokota, head of the study group and a paediatrician at Yokohama City University's Graduate School of Medicine, admits that the study had shortcomings, including a poor definition for the term 'abnormal behaviour'. So in February, at the government's request, Yokota's team launched a larger study, which will trace 10,000 people aged 0–18 years. The team aims to release the results by this autumn.

Experts call for active surveillance of drug safety

Earlier this month, the US Food and Drug Administration (FDA) cautioned Americans that two classes of drugs in widespread use have serious and sometimes life-threatening side effects that warrant new label warnings.

On 9 March, the agency announced that three erythropoietin-type drugs for treating anaemia increase the risk of death, blood clots, strokes and heart attacks, and accelerate the growth of some cancers. Then on 14 March, it asked makers of 13 widely prescribed sedative-hypnotic sleeping pills to toughen label warnings because of reported severe allergic reactions and "complex sleep-related behaviours", including falling asleep while driving.

These warnings, for blockbuster drugs that have been on the market for years, highlight an increasing challenge that confronts the FDA. Side effects may not emerge until new drugs are taken by huge numbers of people. This lesson was driven home by the anti-inflammatory drug Vioxx, which was withdrawn in 2004 after five years on the market, for causing heart attacks and strokes. As a greying America consumes an ever-expanding smorgasbord of medicines, the agency is fighting an uphill battle to catch the next Vioxx before it turns into a fully fledged disaster.

So experts are calling for the FDA to move beyond its current passive system for monitoring the safety of marketed drugs. The Adverse Event Reporting System relies on doctors — if and when it occurs to them — to report suspected side effects to the agency. But doctors filed fewer than 25,000 such reports in 2005, suggesting that the system is capturing only a small fraction of the actual side effects.

"The system in this country for identifying drug safety signals is not nearly as robust as it could be," says Scott Gottlieb, resident fellow at the American Enterprise Institute, a conservative Washington think-tank, and a former deputy commissioner for medical and scientific affairs at the FDA.

Last week, Mark McClellan, a former FDA commissioner, told a Senate committee considering new drug-safety legislation that the system "needs to do better than just seeing the tip of the iceberg of a safety problem after it has already hit us". Health-information

R. FRIEDMAN/CORBIS

Side effects often don't emerge until a drug has been on the market for some time.

Suicides in Japanese teenagers have raised concerns about Tamiflu's safety.

But Yokota's team has itself come under fire recently. Last week, it emerged that Chugai paid for paediatric research and teaching by at least two members of the study group, including Yokota, and Tsuneo Morishima of Okayama University.

Chugai gave ten million yen (around US$85,000) to Yokota between 2001 and 2006 and two million yen to Morishima in 2005, prompting criticisms that the researchers had a conflict of interest. "It's quite natural for the public to imagine any interest may have been involved," says Masayuki Shibuya, vice-president at Tokushima University, which has developed model guidelines to manage conflicts of interest in clinical research.

Morishima, Yokota and their universities have defended the donations, saying that all funding was approved by the university, and that the money was used for unrelated work and so did not affect the results of the flu study. Tamaki Fushimi, head of drug safety at the Japanese health ministry, says that the ministry was not aware of the funding when it appointed the study group, and that it has not yet decided whether to allow those who received money from Chugai to continue with the study. Masanori Fukushima, professor of clinical-trial design and management at the Kyoto University Graduate School of Medicine, is one of those who think that the researchers should be removed. "The ministry should care about ethical and regulatory issues more properly," he says.

But Yokota insists he has done nothing wrong. "Tamiflu is an important drug, but we have no standards to tell doctors who can be prescribed it and who can't," he says. "It is our responsibility as paediatricians to create this measurement." ■

Ichiko Fuyuno

technology for drug safety is "an idea whose time has come", he added.

McClellan and others are pushing the idea of data mining of existing health-record databases as an active surveillance system to pick up early warnings of adverse side effects. By pooling existing databases run by private insurance plans, government agencies and industry, information on well over 100 million people could be studied in something much closer to real time, he says. The FDA could then identify priority questions and the mechanisms for answering them. If such a system had been in place when Vioxx came on the market in 1999, that drug's dangers could have been detected in months rather than years, McClellan contends.

The same thinking is gaining currency in Europe. "There is great potential in the use of these healthcare databases," says Panos Tsintis, who oversees the safety of marketed drugs for the European Medicines Agency (EMEA) in London.

Drug makers are also backing the idea of a centralized active surveillance system, which they see as preferable to a duplication of effort in which each insurer, government agency and company invents its own scheme. Such a system "would provide information we crucially currently don't have", said Ron Krall, senior vice-president and chief medical officer for GlaxoSmithKline at an Institute of Medicine (IOM) forum on 12 March in Washington DC. The FDA is testing active surveillance in a pilot programme tracking the side effects of four unnamed, newly approved medicines. "In general, safety information trickles in and the FDA is passive in analysing it," said Ellis Unger, an official from the agency's drug-review centre, at the IOM forum. "In this programme we grab each new product by the horns. It's very resource-intensive."

Those costs present a major problem to the cash-strapped agency, which has no funds even to upgrade its existing surveillance system. But advocates of active surveillance say the country cannot afford not to invest what would probably be tens of millions of dollars. That kind of money, says Garret FitzGerald, head of pharmacology at the University of Pennsylvania in Philadelphia, "is peanuts when you think about the public health". ■

Meredith Wadman

Upstart forum created for German conferences

Several key players in the prestigious but troubled Dahlem Conferences, which have organized some 90 international, multidisciplinary meetings in Berlin since 1974, have resurfaced in Frankfurt as founders of a new forum. The FIAS forum, based at the Frankfurt Institute for Advanced Studies, is meant to restore "the format and philosophy of the original Dahlem Workshop model," the founders say.

Dahlem was integrated into the Free University Berlin in 1990. Troubles surfaced publicly in early 2005 when several senior scientists threatened to resign from the Dahlem board in protest against the firing of programme director Julia Lupp by Free University administrators (see *Nature* 443, 446; 2005). The protestors also alleged that Free University administrative actions had slowed down publication of manuscripts and undermined Dahlem's independence.

Lupp is now in Frankfurt, as programme director and series editor for the new forum. The forum will organize three or four workshops a year with "the same broad focus of the Dahlem workshops," says Wolf Singer, a director of the Max Planck Institute for Brain Research in Frankfurt, and a driving force behind the new forum.

MIT team calls for carbon storage underground

Carbon from coal-burning power plants must be pumped underground to stop it contributing to climate change, according to a report, *The Future of Coal*, from scientists at the Massachusetts Institute of Technology in Cambridge. The team used a computer model to simulate a world in which carbon emissions come at a price. When the cost to emit a tonne of carbon reaches about US$30, it will begin to make economic sense to pay for it to be sequestered underground, the study found.

L. LEE PHOTOGRAPHY/CORBIS

Several new techniques are being tested to store carbon emissions.

The authors call for the construction of up to five large-scale installations, in different locations with different geologies, to show that carbon sequestering can work well. Current projects, they say, are too small and are not monitored well enough to provide useful information.

Big donor cuts Stanford cash because of oil ads

Gatherings of California's Bing family, the mega-donors to Stanford University, could have more energy now that the son has withdrawn a gift of $2.5 million promised to the university his father supports.

Hollywood movie producer Steve Bing last week withdrew the final instalment of a $25-million pledge to the university. A major backer of alternative energy, he rescinded the loan after citing television ads by oil giant ExxonMobil that hyped its own $100-million research grants to Stanford.

His parents, Helen and Peter Bing, have pledged more than $85 million to Stanford, where Peter Bing served on the board for 31 years, including five as chairman. Steve Bing is now calling on others to halt gifts to Stanford, an aide said.

Microsoft provides cash for synthetic biology

The burgeoning field of synthetic biology has a new supporter: Bill Gates.

On 13 March, Microsoft Research in Richmond, Washington, announced the award of $570,000 to six scientists in the United States and Canada to "stimulate foundational research in synthetic biology and DNA nanotechnology by identifying and addressing the unique computational challenges of these areas". The grants will support projects that take computational approaches to tackle biological problems, such as rational gene, genome and protein design and construction.

One grant will help Herbert Sauro of the University of Washington in Seattle build a software tool to aid the assembly of biological devices. 'Plug-and-play' assembly of such devices has been hampered by a lack of standardized information about each biological part created by scientists. Sauro therefore hopes to use part of his grant to start working towards a consensus on design standards.

Germany counts the cost of climate change

Climate change could cost the German economy up to €800 billion (US$1.1 trillion) by 2050, according to a study by the Berlin-based German Institute for Economic Research (DIW).

The authors modelled the effects of a 4.5 °C warming on all economic sectors. Unstopped warming would cause an average loss of 0.5% in national economic growth per year, they conclude.

Few studies on the economic effect of climate change go down to the level of individual countries. The *Stern Review on the Economics of Climate Change*, published last year, predicts long-term global economic losses of 5–20% if nothing is done to slow global warming (see *Nature* 444, 6–7; 2006). The less pessimistic predictions by the DIW are based on 'damage functions' developed by Richard Tol of Princeton University, who says that the *Stern Review* over-estimates the future costs of global warming.

Changing spots

A. COMPOST/WWF-CANON

The clouded leopard found on Borneo and Sumatra is an elusive, reclusive creature that lives in mountainous rainforest. So perhaps it is not a surprise that scientists have only now realized that this leopard is a different species than the clouded leopard of mainland Southeast Asia.

Genetic tests done at the US National Cancer Institute in Bethesda, Maryland, suggest that Borneo's clouded leopard diverged from the mainland cat some 1.4 million years ago. Comparative studies of skins and furs held in museums also support the idea of separate species, say Andrew Kitchener of the National Museums of Scotland in Edinburgh and his colleagues in a recent paper in *Current Biology*.

In Borneo, the last holdout of the clouded leopard is an area of rainforest the size of Kansas, which government officials last month signed an agreement to protect.

FDA proposes tighter rules on conflicts of interest

After a flurry of criticism for perceived conflicts of interest, the US Food and Drug Administration (FDA) is planning to significantly tighten the rules that govern when financial conflicts should exclude experts from serving on its external advisory committees. The committees are important because the agency nearly always follows their advice on approving drugs and devices and on emerging safety issues.

In its draft policy released on 21 March, the FDA said that individuals will generally be excluded from participating on advisory committees if they have financial interests exceeding $50,000 in the issue being discussed. Experts with financial interests of less than $50,000 might be allowed to participate in discussions without voting, the agency said. The draft proposal is open for public comment until 21 May.

Flu study faces shake-up over industry funding

Japan's health ministry is expected to remove two researchers from its eight-member study group on influenza, because their research in other areas was partly funded by a Japanese distributor of the flu drug Tamiflu.

The study of some 10,000 children will investigate the possible side effects of Tamiflu (oseltamivir) as part of its remit.

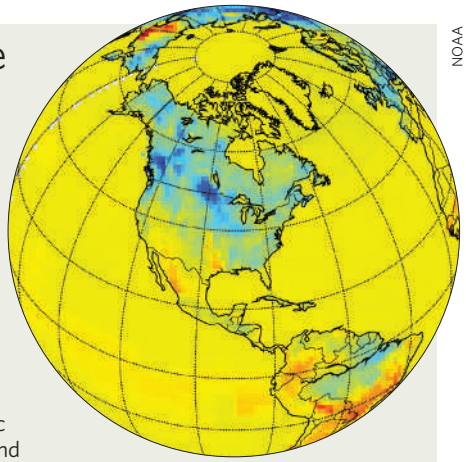
The health ministry last week warned doctors not to give Tamiflu to teenagers, after a number of new reports linked the drug to psychiatric effects such as suicidal tendencies among the age group (see *Nature* doi:10.1038/446358a; 2007). Swiss drug firm Roche, which makes Tamiflu, says that no such causal link has been established, and the World Health Organization says Tamiflu remains the drug of choice for treating people infected with the bird flu virus H5N1.

Shunpei Yokota of Yokohama City University, the study group's leader, and Tsuneo Morishima of Okayama University have received some ¥12 million (US$100,000) between them from Tokyo-based Chugai Pharmaceutical, which distributes products for Roche. Health minister Hakuo Yanagisawa told a parliamentary committee on 23 March that the researchers ought to be excluded from the flu study.

KYODO/NEWSCOM

The Japanese government wants to investigate the possible side effects of flu drug Tamiflu.

On the track of carbon dioxide



NOAA

It's still pretty raw, but an online tool to track carbon dioxide emissions is being set up by the US government. The new CarbonTracker website (http://carbontracker.noaa.gov) is meant to provide a public-friendly view of greenhouse-gas emissions from various sources around the world. The snapshot on the right shows CO_2 uptake for a week in July 2005 — dark blue represents the strongest CO_2 sinks.

So far, though, the data are sparse. Only 60 carbon monitoring sites worldwide are included, with 20 of them in the United States. Project scientists at the National Oceanic and Atmospheric Administration say they hope to add further sites and eventually develop CarbonTracker into a resource for policy-makers and scientists.

UK league table revisits problems caused by drugs

Alcohol and tobacco are better than heroin but worse than cannabis, according to a UK ranking of the dangers of recreational drugs (D. Nutt *et al. Lancet* 369, 1047–1053; 2007).

The new system is an attempt to provide a scientific — if still simplistic — way to compare the social and health tolls taken by recreational drugs. Current British drug laws are shaped by political prejudice as much as by the actual threats posed by the substances, says team member David Nutt of the University of Bristol.

His team asked experts — including psychiatrists specializing in addiction, the police, forensic experts and doctors — to give up to 20 drugs a score in nine subcategories within the larger categories of physical harm, dependence and social harm.

The result? Heroin and cocaine were ranked as the most dangerous, reflecting their status as class A drugs — the most harmful tier of Britain's three-category system. But ecstasy, another class A drug, finished eighteenth in their list — below commercial solvents and anabolic steroids.

Budget gives Canadian science a cash injection

Canada's science infrastructure got a boost last week with the release of the country's budget plan for 2007.

The two main science granting agencies — the Natural Sciences and Engineering Research Council and the Canadian Institutes of Health Research — will each get an extra Can$37 million (US$32 million), raising their combined budgets to roughly Can$1.4 billion. A further Can$510 million is allocated to modernize the research infrastructure at universities and other research institutions. And Genome Canada will receive an extra Can$100 million for grants and regional genome centres.

Another winner, to the tune of Can$50 million, is the Perimeter Institute for Theoretical Physics in Waterloo, Ontario. Seven other targeted institutes, including some focusing on neurology and sustainable energy, will share an extra Can$105 million.

SpaceX rocket burns up after missing orbit

The privately financed Falcon 1 launch vehicle reached an altitude of 300 kilometres last week before developing problems and burning up on re-entry into Earth's atmosphere. But space-industry experts say the test should be considered a success.

Falcon 1 was developed by the California-based company SpaceX as a rapid way of getting satellites into orbit. Space launches usually take months of planning, but industry observers say Falcon 1 should be turned around quickly because SpaceX has used a simple design that the firm says will eventually be operated by as few as 15 staff.

SpaceX is investigating why the Falcon 1 vehicle developed a rolling motion during the flight, which caused its engines to shut down, about 6 minutes after launching from the Marshall Islands on 20 March. The company is still aiming for another launch later this year to put a US Department of Defense satellite into orbit.

Correction

Our News in Brief story 'Upstart forum created for German conferences' (*Nature* **446,** 360; 2007) contained an incorrect reference to an earlier story on the topic. The correct reference is *Nature* **433,** 446 (2005).

BUSINESS

There goes Texas

A big deal in a big, conservative American state shows that energy utilities can no longer ignore CO_2 emissions. **Emma Marris** reports.

The state of Texas has sometimes chafed at being a mere state, and not a country unto itself. It likes to do things its own way. And the way it is doing energy, to judge from a ground-breaking takeover deal announced last month, is very different now from how it would have been just a year ago.

The proposed buy-out of Texas's largest electrical utility, TXU, by a private-equity investment group has been endorsed by two of the leading environmental groups in the United States. It contains elements which promise to yield a major slow-down in the growth of carbon dioxide emissions in this most energy-addicted of states.

Texans love their way of life, but all those air-conditioning systems, trucks and miles of open road make the state a monster emitter. Carbon dioxide emissions per capita are high (see chart), with total estimated emissions of 700 million tonnes in 2003.

The Texan setting — with its emissions, conservative politics and love of all things big — makes a striking venue for a deal which, analysts say, sets the pace for how businesses are beginning to take action to restrain carbon emissions.

Dallas-based TXU is, subject to state regulatory approval and the agreement of its shareholders, set to be purchased by a group of private investors led by Texas Pacific Group of Fort Worth, Texas, and Kohlberg Kravis Roberts & Co in New York for $45 billion. In an arrangement unprecedented in the US energy sector, the deal was endorsed by two leading environmental groups — the pro-business Environmental Defense and the more left-leaning Natural Resources Defense Council, both based in New York. Participants say it took a 17-hour negotiation in a San Francisco hotel to secure the endorsement.

To get it, the buy-out group pledged to abandon eight out of 11 new coal-fired power stations that TXU had been planning to build (although the present management might have put these on hold in any case). The group says the company will spend $400 million over five years encouraging TXU's customers to use less power. It will also join other utilities, such as the San Francisco-based Pacific Gas and Electricity Company, in publicly backing a federal mandatory cap-and-trade scheme for carbon emissions. Commitments were made to reduce carbon emissions to 1990 levels by 2020. And the company has pledged not to go out of Texas to states such as Virginia or Pennsylvania, where carbon emissions may be more lightly regulated, and build power plants there.

William Reilly, who ran the Environmental Protection Agency under President George Bush senior and who is a former president of environmental charity WWF, will sit on the board of the new company.

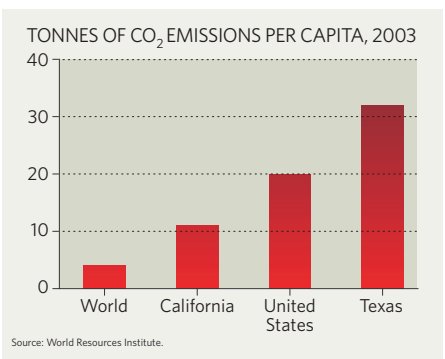
Continued objections

Not everyone is happy with the proposed deal. Linda Miller, the mayor of Dallas, and several environmental groups have said that they will continue to oppose the three coal-fired power plants permitted by the agreement. "The fact that Environmental Defense gives its blessing doesn't mean that more radical groups are going to toe the line," says Rob Wilder, whose San Diego consultancy, WilderShares, follows the clean-energy sector. "It doesn't shield them completely from opposition, but I think it will be seen as being helpful."

The arrangement does appear to have shielded the buyers from criticism they might have expected for making a heavily leveraged buy-out that will take TXU from being a publicly listed company to a private one. Private buy-outs have been attacked for reducing public accountability, piling up debt and encouraging the pursuit of fast returns, sometimes at the expense of customers and employees.

TXU's buyers say that they are with the company for the long haul, although they do plan to split the utility into three components, dealing with power generation, distribution and sale.

Another criticism has been that none of the environmental commitments is legally binding. But Jim Marston, head of Environmental Defense in Texas, says the deal was preferable to continued court wrangling with TXU over its power-station plans. "Look," he explains, "these guys said 'We have a business model that can make money in a carbon-constrained world.' Now that was music to my ears."



Going green may, in fact, have been the only option for TXU, trapped between burgeoning electricity demand from its customers and resistance to its plans for new capacity. Two-thirds of its electricity currently comes from natural gas, which has almost doubled in price since 2000. An extensive strip of brown coal, or lignite, that runs through the middle of the state seemed like the best way to augment electricity supply cheaply.

TXU's announcement early this year of plans to build 11 coal-fired plants blew up a storm of opposition. Democrats in Washington warned publicly that all emissions from new coal-fired capacity in Texas would be treated as additional under a federal cap-and-trade scheme. The Texas legislature started work on a bill to freeze new coal-plant construction for two years. City mayors from Dallas, Houston and elsewhere were fighting the plant permits in the courts, with *pro bono* help from Stephen Susman, a hotshot Houston lawyer.

The message was clear: carbon emissions were becoming a major problem for TXU. The group of investors reached the agreement with this in mind, says Reilly. "If it was not possi-

D. J. PHILLIP/AP

Coal-fired power plants such as TXU's Big Brown get the thumbs-down in Texas.

ble to get the environmentalists on board, we weren't going to do the deal." Reilly says the increasing need for power in Texas will be met with a mix of clean and less clean technologies and demand-side management.

A report prepared in January by energy-efficiency consultancy Optimal Energy of Bristol, Vermont, in response to the original 11-plant plan, suggests that tough, California-style efficiency programmes, plus combined heat and power schemes, could keep pace with Texan demand for electricity, which is expected to rise by about 2% each year.

Other bidders for TXU have until 16 April to come forward — although a recent fall-off in the company's share price suggests that the market does not expect them to. Daniele Seitz, analyst at New York investment bank Dahlman Rose, says that this deal is for big fish only. Any buyer would also be obliged to replicate environmental aspects of the deal, other analysts say. This shows clearly that carbon emissions can no longer be ignored — even in Texas.

Outside observers caution, however, that any plan to meet growing electricity demand without resorting to coal has yet to deliver. "This hasn't been stress tested," says Michael Liebreich, head of New Energy Finance, a London-based energy consultancy. "If there are blackouts, all bets are off." ■

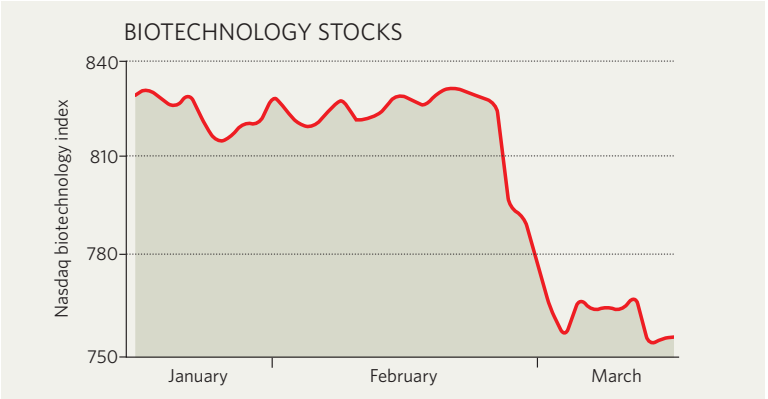
IN BRIEF

OUT OF THAILAND As tension mounts over drug-pricing policy, pharmaceutical company Abbott says it will not be launching seven new drugs in Thailand, including treatments for HIV and bacterial infections. The move follows the country's decision to issue 'compulsory licences' that allow it to make or import cheap versions of several patented drugs. One of these is Kaletra, an anti-HIV drug from Abbott. The company says it decided to withhold the new drugs on the grounds that it can no longer defend its patents in Thailand.

SCOT FREE Charges brought against senior managers of computer company Hewlett-Packard for spying on their own employees and board members have been dropped. In a 14 March settlement at the Santa Clara Country Superior Court between the judge and the California state prosecutors, all charges against Patricia Dunn, the former company chairman who had to resign last year over the scandal, were dismissed, and charges against the other four defendants have been reduced to misdemeanours. The charges rocked Hewlett-Packard after Dunn and her colleagues launched a no-holds-barred investigation into the leaking of details of company board meetings to the media.

ILLUMINATION DELAYED Attendees at the American College of Cardiology meeting in New Orleans next week won't see the much-anticipated results of the study that forced Pfizer to pull the plug on its next-generation cholesterol drug torcetrapib (see *Nature* **445,** 13; 2007), the company says. The detailed results were to have been released at the meeting, but are taking longer to analyse than expected. Before the 15,000-subject trial was stopped last December, deaths among patients taking torcetrapib were significantly higher than among those who were not taking the drug.

MARKET WATCH



This week Wood Mackenzie, an Edinburgh-based research and consulting firm, reviews recent trends in biotechnology stocks.

From mid-January, the Nasdaq Biotechnology Index set out on a gradual upward trajectory — only to be interrupted by a sharp drop in late February. This fall wasn't biotechnology-specific, however — it reflected the global market correction at the end of that month.

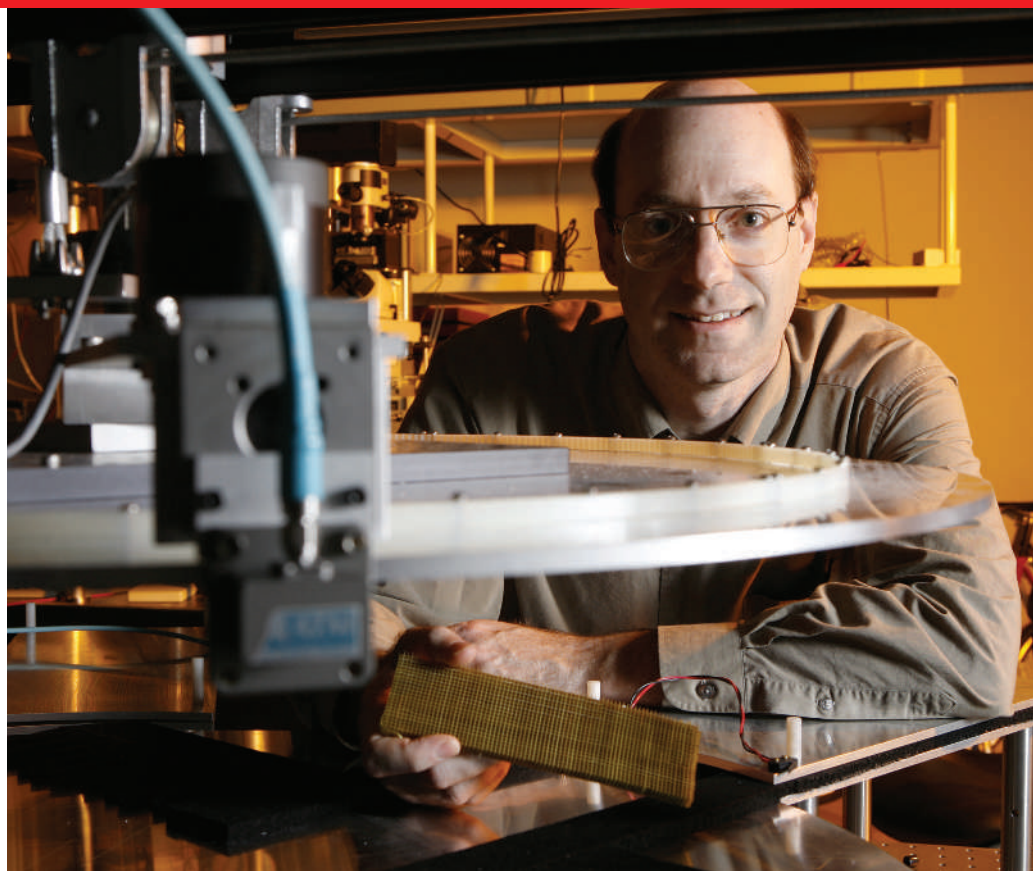
The biggest mover in the index over the past two months is California-based Onyx, whose stock price more than doubled in the week starting on 12 February. These gains — which survived the correction — came in response to strong data for its oncology drug, Nexavar, in late-stage clinical trials for the treatment of liver cancer. Onyx's partner for Nexavar is the German drug firm Bayer Schering, which already markets the drug for renal cancer treatment in the United States and the European Union.

Other gains were triggered by robust financial performance reported by a small number of biotech firms. Gilead Sciences, also of California, posted a 48% year-on-year rise in fourth-quarter revenues, underpinned by a surge in sales of its HIV drugs. The market liked Gilead's forecast that total sales could reach $3.5 billion in 2007, up 35% from last year: the company's shares duly hit their highest-ever level on 21 February.

The biotechnology index underperformed broader indices last year and is now faced with the challenge of recovering from the February dip. The performances of these Californian companies provide early indications that it may yet manage to do so. ■

Unexpected tricks of the light

The buzz over invisibility cloaks is fun — while it lasts. But metamaterials are likely to transform optics through more mundane applications. **Katharine Sanderson** reports.



David Smith was the first to demonstrate how a metamaterial could have a negative refractive index.

J. WALLACE/DUKE PHOTOGRAPHY

There is a world where objects can be made invisible, where light can be bent the wrong way, and where images of incredibly small objects can be brought into sharp focus by a superlens. That magical world doesn't sound very real, and it isn't. It exists mostly in the minds of theoretical physicists. But under the guise of metamaterials, some of these weird properties are making it into the real world, and beginning to attract commercial interest that goes beyond boyish excitement about 'invisibility cloaks'.

In reality, metamaterials applications are likely to be much more mundane, says Michael Wiltshire, a researcher at Imperial College London. Wiltshire predicts that metamaterials will find their first uses in antennas for telecommunication or in biological imaging. It's a far cry from last year's headlines about Harry Potter and magic cloaks but a first step to a world of new optical tricks.

Metamaterials are engineered to have features that are about the same size as, and usually smaller than, the wavelength of electromagnetic radiation they are being used to manipulate. The features are often small metallic wires and coils, which together manipulate the electrical and magnetic components of electromagnetic waves in seemingly unnatural ways. It is these tiny features of their design, rather than the atomic structure of the material itself, that determine its properties.

"Metamaterials are an answer searching for a question." — Nathan Myhrvold

The field got going in 2000 when John Pendry at Imperial College suggested that metamaterials could be used to make a 'superlens'[1] that could image objects smaller than the wavelength of the incoming light. Making a lens that could bring such tiny features into focus would require a material with a negative refractive index, one that would actually bend incoming light the wrong way. This was achieved in 2001, although not for visible light, when David Smith at Duke University in Durham, North Carolina, made a two-dimensional metamaterial from interlocking units of fibreglass board patterned on one side with a copper strip and on the other with split-ring resonators[2].

But the public's imagination was really caught in May 2006 when Pendry suggested that, theoretically, metamaterials could be used to make an invisibility cloak. Again, Smith's team followed just five months later with a practical demonstration of invisibility. Smith's cloak used concentric rings of fibreglass imprinted with copper wires (see opposite). This contraption, when placed around an object, directed microwaves around that object, much as a stick dipped in a stream interrupts the flow of water. When the microwaves met up behind the object they behaved as if the object hadn't been there[3].

This cloaking prototype only works in two dimensions, and for a narrow band of wavelengths, but it was a proof of principle. With this demonstration came promises of the ultimate camouflage — in short, applications that would have military agencies salivating.

According to Pendry, the first metamaterial was developed by the US military in the 1950s, and its interest in the topic has never died. "The primary interest for sure is from the military," says Smith, who received most of his research funding from the military. In 2001, the US Defense Advanced Research Projects Agency (DARPA) set up a metamaterials research programme. In total, it has invested US$40 million over six years — enough for a young field to get going but small change for an agency whose annual budget is several billion dollars.

"Metamaterial developments so far are exciting and promising," a DARPA spokesperson says, "but they still require significant development to qualify for 'real world' applications." Not even Smith thinks that a fully functional invisibility cloak will become a reality. "We won't be able to cloak a fighter jet," says Smith. "I don't know how large a structure we could eventually cloak. Right now, we are trying to play around with the parameters to implement a full cloak, but it is much more difficult."

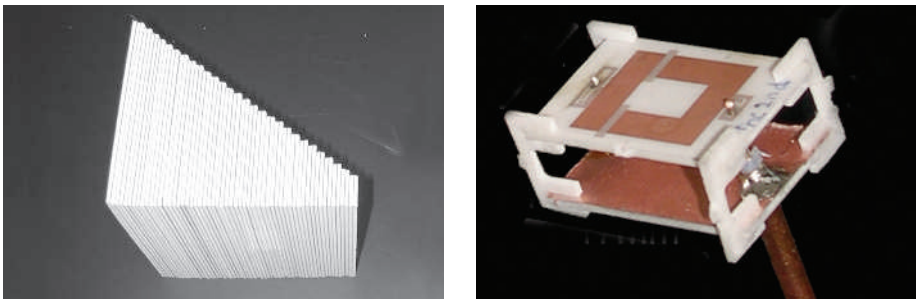
Mind your Ps

Pendry was working on radar-absorbing materials for a military contractor when he came across the work of Victor Veselago, a physicist at the Russian Academy of Sciences in Moscow. Metamaterials take advantage of two fundamental properties of materials — their electric permittivity and magnetic permeability, which are the amounts by which the electric and magnetic parts of a wave interact with a material. Most, but not all, materials have positive values for these two parameters. And in the 1960s, Veselago expanded previous work to study what happens when both the permittivity and the permeability are negative. This leads to a negative refractive index — the amount by which light is bent as it passes from one medium into another — and as Veselago showed, to new optical properties. It was Pendry who realized that with metamaterials Veselago's ideas could become reality.

DARPA's initial interest was in applications beyond optics, such as magnets for more pow-

erful motors, radar applications, and textured surfaces that could control a material's heat-handling capabilities. DARPA now thinks, and Wiltshire agrees, that the most feasible advances are likely to be ultra-high-frequency antennas for possible radar applications.

DARPA won't reveal much about its radar work, but others are pursuing commercial applications of metamaterials for radio-frequency antennas. Conventional antennas need to be half as high as the wavelength of the radio waves, which can be many metres for low radio frequencies. More compact antennas are needed for the next generation of hand-held devices such as mobile phones and laptops.



J. B. PENDRY; G. ELEFTHERIADES

Negatively refracting materials (left) are being used in devices such as the compact antenna (right).

Fast and small

To reduce the wavelength (and the size of antennas), one option is to increase the frequency; at 30 gigahertz the wavelength is one centimetre. But these higher-frequency waves are blocked more easily by obstacles, making them impractical for networks in urban areas. In 'normal' materials the wavelength is inversely proportional to the frequency of the wavelength, but in a negative-index antenna, something strange happens — the apparent wavelength decreases as the frequency decreases. "This is quite the opposite of what happens in conventional materials," says George Eleftheriades, an electrical engineer at the University of Toronto in Canada.

According to Eleftheriades, by exploiting this property, the height of the antenna could be just one-thirtieth of the wavelength, making compact millimetre-sized devices a real possibility (see above). "Now, the wavelength reduces as the frequency is reduced," says Eleftheriades, "hence short antennas can be constructed even at the low frequencies (1–5 gigahertz) at which most wireless telecom systems work."

Developing metamaterials for antennas has allowed Eleftheriades to bag a commercial contract with Nortel Networks in Toronto, and a project funded jointly by Nortel and the Canadian government to the tune of $400,000. He also has a patent on his antenna design.

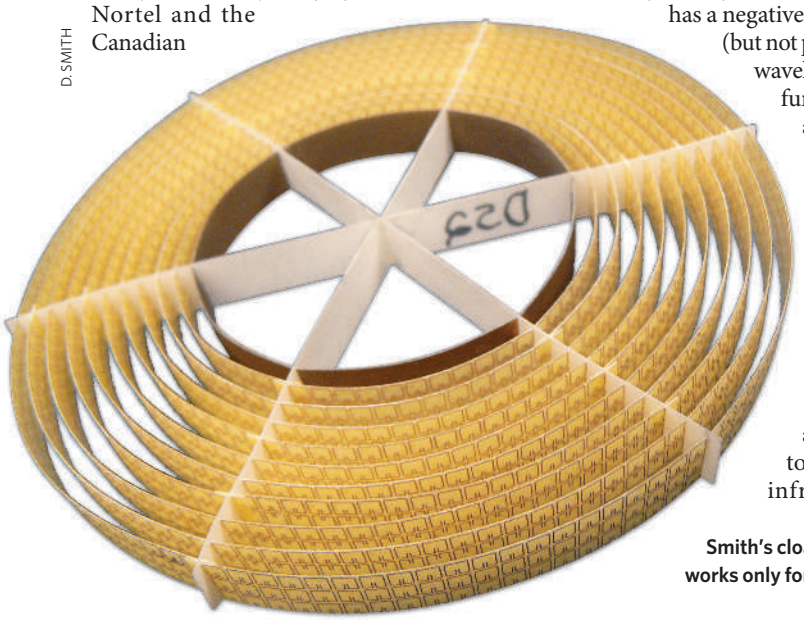
Beyond antennas, "the optical range is perhaps the most practically important for applications," says Pendry. Optical communication devices, such as fibre-optic cables, and biological imaging would be enhanced if light at optical wavelengths could be manipulated in new ways. But shrinking a metamaterial's features to match these wavelengths is more challenging: the material would need features that control multiple wavelengths of light simultaneously — and that would involve some very small and complicated engineering.

By making optical devices, physicists hope to beat the 'diffraction limit' and so achieve a superlens. For features smaller than roughly half the wavelength of light, diffraction in a normal lens distorts the image. A negative-index material that could focus visible light from subwavelength objects should produce distortion-free images. This would revolutionize biological imaging, Pendry argues, by bringing the details of cells and cellular processes into sharp focus.

"The optical range is perhaps the most practically important for applications."
— John Pendry

In 2005, two teams from the United States and New Zealand showed that a simple superlens could be created for ultraviolet wavelengths from very thin layers of silver[4,5]. Silver naturally has a negative electric permittivity (but not permeability) at these wavelengths and is able to function as a superlens as long as the distance between the object and lens is much less than the wavelength used. These experiments achieved only modest gains over the diffraction limit, but such superlenses are already being used to improve near-field infrared microscopy[6].

D. SMITH



Smith's cloaking device currently works only for microwaves.

Smith admits that losses and imperfections in the metamaterial will limit the resolution that can ultimately be obtained, but insists, "The superlens as a concept is sound."

"It is a separate question as to whether any of these metamaterials technologies will be useful or commercial — that will depend on the competition," Smith says. He argues that metamaterials ideas are already influencing the next generation of engineers: "So these concepts may seep into the techniques and bags of tricks that engineers employ without necessarily being identified as 'metamaterials'. It could be a very quiet revolution."

For example, the aerospace company Boeing in Seattle, Washington, has explored the use of metamaterials with negative indices as lenses, again funded mainly by DARPA. Lenses are used in satellites to help focus microwave signals onto antennas, enhancing their detection from the ground. Using a negative-index lens in a communications satellite would enhance performance while reducing the weight of the lenses. This is a simple, albeit niche, application of a metamaterial, says Wiltshire. "What we're exploiting is nothing sexy in the metamaterial — it's mundane — the lens is lighter," he explains.

Nathan Myhrvold, chief executive of Intellectual Ventures in Bellevue, Washington, and formerly Microsoft's first chief technology officer, claims to have more patents for metamaterials than anyone. "Dozens of them," he says. The issue for metamaterial applications is not so much when, but rather where they will be used, he says: "At the moment metamaterials are an answer searching for a question — certainly from a commercial standpoint." Like many in the field, he has faith that the technological breakthroughs will come. ■

Katharine Sanderson is a reporter for *Nature* based in London.

1. Pendry, J. B. *Phys. Rev. Lett.* **85,** 3966–3969 (2000).
2. Shelby, R. A., Smith, D. R. & Schult, S. *Science* **292,** 77–79 (2001).
3. Schurig, D. *et al. Science* **314,** 977–980 (2006).
4. Fang, N., Lee, H., Sun, C. & Zhang, X. *Science* **308,** 534–537 (2005).
5. Melville, D. *et al. Opt. Express* **13,** 2127 (2005).
6. Taubner, T., Korobkin, D., Urzhumov, Y., Shvets, G. & Hillenbrand, R. *Science* **313,** 1595 (2006).

PA/EMPICS

Lives in limbo

Many Iraqi academics have escaped death threats only to find that their qualifications are obsolete and immigration authorities are unsympathetic. **Jim Giles** hears their stories.

Some 20 years ago, Ali Althamir led a comfortable life in Iraq. As head of a university department and an expert in computing, he was part of a well-heeled middle class. He worked at one of the better institutions in the Middle East. And before sanctions crippled the regime of Saddam Hussein, he was one of many researchers who had money for their studies and could sometimes travel to foreign conferences.

It is a far cry from the life Althamir lives now. Since arriving in Britain to seek asylum in 2003, his fingerprints have been placed on file at the police station and he reports there once a month. Filling in the time between visits is difficult because he cannot now legally be paid to clean a lecture hall, let alone speak in one. When I thank him for taking time to meet me, he tells me not to worry. "I have nothing to do," he says.

It's a drab January day when I meet Althamir and other Iraqi scientists and clinicians at an office in the London South Bank University. They have travelled across the capital to meet me, from cheap, rented accommodation paid for by social services and charitable hand-outs. With Althamir are an expert in radar technologies, a civil engineer and a former lecturer at Baghdad's most prestigious medical school. Between them, they share well over half a century of research experience, the painful decision to flee their country — and little idea of what the future holds.

The emotions they express are predictable: anger that a nation that invaded their country is now treating them so badly; frustration at the unfathomable British immigration laws; fear for family and colleagues they have left behind. They have fled death threats only to run into an unwelcoming immigration system and an indifferent academic community that often does not recognize their experience.

They did not know that Britain's growing immigrant populations have been a source of acute political tension for decades, tension that has led to tight restrictions on who can enter and complex rules to control the country's gateway. Nor did they know that experience such as being a head of department would do little to help them find work. Learning about these obstacles beforehand is not always a priority, or even a possibility.

It is a physical and emotional journey that has been shared by many since the invasion of Iraq in 2003. The United Nations estimates that half a million Iraqis left their country last year alone. Some of the academics who want to continue their research head to Britain, which has strong historical links with Iraq and is where many Iraqis studied at the start of their careers. They also go to the United States because of its strong reputation for research and funding. No hard figures are available, but several thousand academics are thought to have left Iraq since the invasion, of which at least a hundred have entered Britain and the United States.

Psychiatrist and researcher Ali Omar, who asked that his real name not be used, tells a story that many Iraqi academics will recognize. His life entered a radically new chapter on 9 April 2003, the day American troops

famously toppled the statue of Saddam Hussein in central Baghdad. The looting began almost immediately. "Everything went," Omar recalls, "doors, air conditioning, lights."

Omar found his clinic destroyed and spent much of the time afterwards building up a network of psychiatrists in Iraq to help those harmed by the war. He later organized a conference that attracted media attention — it should have been a triumph, but immediately afterwards relatives called. The publicity, they said, would make him a target of the militias and criminal gangs that were kidnapping and killing academics. They were right.

MIRRORPIX/GETTY IMAGES

The toppling of Saddam Hussein's statue marked the start of major upheaval for many Iraqi academics.

Targeted groups

Around 300 academics have been killed since the invasion, according to human-rights groups and media reports. No one is sure whether Shia, Sunni, Baath or anti-Baath groups are to blame, or exactly why academics are being targeted. Some researchers blame fundamentalist Islamic groups, who are accused of wanting to destroy Iraq's middle class to establish religious rule. In addition, criminal gangs could be the cause, seeking to make money through kidnapping affluent professors and researchers.

In Omar's case, trouble began when he was away at conferences in 2005. Neighbours reported that strangers had been inspecting his family home. In April 2006, he received an anonymous phone call and the warning to "leave or die". Within an hour, he and his family had fled to nearby relatives — by July they were in Jordan.

Omar arrived in Britain later that month on a six-month visitor visa to attend a conference at London's Royal College of Psychiatrists. Once in the country, he says that he arranged to speak to a barrister through a friend. The advice made him hopeful that he could qualify for asylum within a few months because of the death threats against him, and that, if successful, his family would be able to join him.

Immigration lawyers say that the advice from the barrister was flawed. It is now eight months since Omar's claim went to the Home Office and no decision has been made. The delay should have been expected, say lawyers, given the backlog of 6,000 asylum applications that the Home Office says has built up, in part because of underfunding and the complex immigration rules. And because Omar arrived on a visitor visa, he is not allowed to work and is banned from receiving financial benefits. The money he brought with him, generated by selling his possessions in Iraq, has now run out. Yet if he were to leave Britain now he would forfeit his application. "I'm desperate," he says. "I don't know what to do."

Support structures

For academics like Omar, the only means of support is charity, and in his case the Council for Assisting Refugee Academics (CARA) has come to the rescue. The London-based organization, conceived in 1933 to help academics forced out by the Nazis, has provided accommodation and maintenance grants and flown his family to safety in Cairo.

Organizations such as CARA are keen to help academic refugees because "they represent the core of their country's scientific and cultural capital and have a key role to play in the rebuilding of their country," says Kate Robertson, CARA's deputy executive secretary. "As educators and independent thinkers, academics are seen to be particularly influential," she says. Preserving such capital also matters on a global scale. Of the thousands of academics helped by CARA in the 1930s, sixteen went on to become Nobel prize winners.

"Academic refugees have a key role to play in the rebuilding of their country."
— Kate Robertson

Like Omar, many Iraqi academics are unaware of the minefield of bureaucracy they will face to stay permanently or to work in Britain. Some could apply to the Highly Skilled Migrant scheme, which provides entry for well qualified individuals. But they then need to produce recent payslips and tax returns to show that they have been employed in their field, and these documents are often not issued in Iraq. Another route of entry is to obtain a work permit through a sponsoring institution before entering Britain. The institution must prove that it has tried and failed to find a better qualified candidate from within Europe, which is difficult and often requires an inside champion for the applicant.

Political barriers

When it comes to gaining asylum in Britain, there could be an additional political impediment, says Chris Randall, an immigration lawyer with Bates Wells and Braithwaite in London, who has advised CARA. Ninety per cent of the Iraqi asylum decisions announced in the past year have been refusals, and Randall thinks that this is partly because British authorities are reluctant to admit that the situation in Iraq is bad enough to warrant asylum. "The government wants to portray the idea that Iraq is getting better," he says. Those waiting to hear are therefore understandably nervous. The academics who asked that their names be changed were fearful of reprisals against relatives back home, but were also worried about antagonizing the UK immigration authorities. The Home Office denies any political bias against Iraqi asylum seekers, saying that every case is taken on its merits.

Even if Iraqi academics overcome the legal obstacles, they frequently struggle to find a job. Sabrine Gilel, once a dentistry researcher at Baghdad Medical College, left Iraq before the invasion. High-ranking party officials in Saddam Hussein's Baath party would send their children to her college in the expectation

that they would be awarded a qualification, even if they hadn't earned one. Gilel refused to wave students through exams and this, together with a family link to a relative who had opposed Saddam Hussein, marked her out.

The breaking point for Gilel, who also asked for her name to be changed in this article, occurred in 1994. United Nations weapons inspectors were trying to discover the extent of the country's weapons facilities. During a break one day, Gilel and a colleague spotted men carrying barrels through the school. "We laughed. We said: these are the chemicals, here they are in our school." Gilel does not know who overheard, but by the next day her colleague had vanished. A couple of weeks later, after being warned by a friendly security official that she was also being targeted, she fled to Jordan.

K. MOHAMMED/AP

Many Iraqi academics have been targeted because of their position in society.

Transferable skills

When she arrived in Britain in 1999, following a stint at a north African university, Gilel wanted to take up her research in dentistry again but found it impossible to get a research position or grants. Initially she did not have full residency status, which was enough to deter some institutions from employing her. And academic research was impossible because the General Medical Council, which oversees British doctors, did not recognize her qualifications and would not allow her to perform studies that involved patients. "All my experience has gone down the drain," she says. Even to practise as a dentist in Britain she had to study for two years and requalify, which she finally achieved last November.

> **"The government wants to portray the idea that Iraq is getting better."**
> — Chris Randall

Even with extensive experience back home, Iraqi academics find it difficult to break into British universities. Many have no connections in Britain, so they try writing fruitlessly to whoever is listed as a contact on a university website. They also have language barriers and are at a disadvantage because they do not understand how the British funding system works — problems that CARA is tackling by setting up a network of university contacts that can help to advise Iraqi academics.

Another harsh reality is that outside Iraq researchers may not have the skills, reputation or publications they need to compete in academia. After Iraq invaded Kuwait in 1990, Iraqi universities were isolated from the rest of the world by sanctions. For many academics, journal articles could be obtained only by sending someone to photocopy papers from libraries in Jordan. So unless they were able to spend time in foreign labs, even the most talented researchers would have fallen behind colleagues abroad. "During sanctions the level of knowledge moved on and left quite a lot behind," says Robertson.

Hatem Al-Delaimi is the author of three textbooks and has more than 20 years engineering experience — but now finds himself desperately trying to update his skills. After fleeing Iraq for Britain in 2002, he sent off countless application forms and never received a single reply, let alone a job offer. But he was able to get unpaid work on a robotics project at Kings College London, and CARA has provided £3,100 to cover around a year of research costs. It's a breakthrough, but perhaps only a temporary one. When the money from CARA runs out, Al-Delaimi will once again need to convince universities to take him on. But by then he hopes to have at least two publications.

In the United States, another major destination for fleeing Iraqi researchers, immigrants find the situation slightly easier. They can still struggle to find an opening, says Rob Quinn, director of the Scholars at Risk Network, a New York-based organization that defends the human rights of academics. But neither Quinn nor others working with Iraqis in the United States report the level of difficulty experienced in Britain.

S. MIZBAN/AP

Iraqi researchers fled their labs for foreign ones.

One reason is that the US visa system is more flexible than that in Britain. Iraqis often go to the United States to work on temporary visas, for example, and can generally get those visas repeatedly renewed, provided the institution they work for backs their application. (No equivalent visa exists in Britain, says Randall.) The Scholar Rescue Fund, based at the Institute of International Education in New York, is currently funding around 20 Iraqi academics on such visas.

More funds

Once in the country, those academics must still convince institutions that they are up to scratch and worthy of employment. But this also seems to be easier, partly because a little more funding is available. American universities often have more flexible budgets than their UK counterparts and can find the money for a one-year contract, particularly if part of it is provided by organizations like the Scholar Rescue Fund.

The Home Office is talking about changing the country's immigration system and moving towards a system that may look more like that of the United States: one that places more emphasis on admitting those with appropriate skills rather than simply controlling numbers. Lawyers who represent immigrants say they will wait to see details before being convinced.

Such changes could take years to materialize — and offer little solace to Iraqis already in Britain, such as computer scientist Althamir. But in the weeks after meeting Althamir, I learn that he has been offered work with a team working on artificial intelligence at a British university. As an asylum seeker, he cannot be paid — but he can just about survive on state benefits and the £2,000 that CARA will provide in research funds. It is a step back into research, although it seems a far cry from what a former head of department might hope for.

But when I ask Althamir how he feels about going back to work, he corrects me: "I am so happy to join the research team; at least I can catch up with up-to-date research. But it is not going back to work," he adds. "I understand 'work' as a paid job." ■

Jim Giles is a reporter for *Nature* in London.

J. MOORE/AP

The 31st Combat Support Hospital in Baghdad is considered the busiest US combat hospital in the world.

The war against wounds

The US military is getting a lot of flak for the way it treats wounded soldiers returning from Iraq. **Emma Marris** reports on the advances in medical care that are helping to bring them home.

Many numbers can be used to tell the stories of war. Almost 3,500 Americans dead in Iraq and Afghanistan. Some 24,000 injured servicemen and women. Other numbers are subject to their own skirmishes: estimates of the number of civilian Iraqi dead are hotly contested and range from tens of thousands to half a million. Every number tells a story.

A little-known number is that since 2001, the wars in Iraq and Afghanistan have had a US military case fatality rate — the percentage of injured soldiers who die — of 9.4. That is, of 100 people wounded, 9.4 of them will die, either instantly or later, from their wounds. This compares with case fatality rates of 15.8 in Vietnam and 19.1 in the Second World War[1]. The reasons for the decline in the current conflicts are better armour, logistics and medical tools.

Battlefield medicine has often led to advances in trauma care. Florence Nightingale's sanitary reforms in hospitals during the Crimean War are a classic example. In the 1950s Korean War, improved resuscitation of the wounded and rapid air evacuation to Mobile Army Surgical Hospitals — the famous MASH units — reduced the US hospital mortality rate by 24% compared with the Second World War[2].

John Holcomb, commander of the US Army Institute of Surgical Research in Fort Sam Huston, Texas, summarizes the advances in the current conflicts as "training, protection and devices". The devices that he and other trauma surgeons point to are better body armour, redesigned tourniquets and advanced clotting aids and transfusion fluids. But all the military personnel interviewed for this story stressed the importance of training and preparation: most surgeons sent to Iraq and Afghanistan are first required to spend time in a civilian trauma centre.

Teasing apart the actual contributions of individual advances to the lower body count is a much greater challenge. Many data are classified or unavailable. Perhaps the most scientific of all the wartime advances in Iraq is the creation of the Joint Theater Trauma Registry. For the first time, solid data are being collected for trauma cases from injury through to recovery. "We are able to get a sense about what kind of improvements can be made in the body armour," explains Paul Cordts, who runs the office that sets policy at the office of the US Army Surgeon General. "We can see what kind of body armour the casualty was wearing, what damage there was to the vehicles, and what happened to the casualties."

"Getting data from the fog of war isn't easy." — Ray Huey

This data collection and analysis is intended to guide future improvements in medical training and equipment, but also to investigate the effect of procedural changes made during the current conflicts. "For all these wars — and we have had a lot of wars — we have never done any thorough data collection," says Basil Pruitt, editor of the *Journal of Trauma*. "Now they can really identify the outcomes of these patients and figure out what is particularly beneficial."

For Pruitt, the steady improvement in survival continues a historical trend: "The survival from serious wounds has increased as the time between injury and arrival at a definitive hospital has decreased. That started in the First World War really." The Korean war took this further by making rapid helicopter transfers to surgical units routine practice. Now, he says, such units are better equipped: "They have a whole panoply of specialists. If you have a head injury, you do better now than you used to do because you have neurosurgeons."

Military medicine often begins just seconds after a wounding, because all soldiers and marines now carry one-handed tourniquets and clotting agents with them to help slow down blood loss. The main agents used are QuikClot, which comes as a sponge or as

powder, and HemCon, which is in bandage form. No one knows exactly how many lives they save: "Getting data from the fog of war isn't easy," says Raymond Huey, chief executive of Z-Medica in Wallingford, Connecticut, the company that makes QuikClot.

Far-forward thinking

QuikClot powder is made of porous minerals called zeolites. The story is that inventor Frank Hursey, who was working with zeolites as sieves to separate gases, cut himself shaving and applied it to his face on a whim. How it works is still unclear, although it has been approved for clinical use. "There is a whole lot of surface chemistry," says Huey. The product also includes calcium ions, catalysts for the body's clotting process. It is carried by every marine and by members of a number of other forces.

HemCon bandages rely on chitin molecules, which are obtained from shrimp shells. The chitin is extremely sticky and glues the bandage to the flesh "like super duct tape", according to Mike Zoormajian, product manager at HemCon Medical Technologies in Portland, Oregon. Then, the positively charged chitin attracts and clumps together the negatively charged blood cells. Zoormajian says that the product was developed after casualties in Somalia bled to death in the streets because no medics could reach them. Since 2005, every US soldier serving in the Middle East carries at least one of these bandages. Zoormajian thinks that the bandages have saved well over 100 lives.

Once a medic gets to a wounded soldier, the solider enters a rapid and organized system.

"Survival will increase hugely when you can get blood or something like it on the battlefield." — Richard Jadick

The casualty is stabilized far-forward — that is, right up in amongst the action — and then sent to a nearby aid station, where they are patched up further and sent to a Combat Support Hospital. From there, they might be flown to Landstuhl Regional Medical Center in Germany or on to Walter Reed Army Medical Center in Washington DC. All this often happens inside a week. Speed really does save people with battlefield injuries. The alternative can be bleeding to death.

Saving lives

For Richard Jadick, a surgeon deep in the heart of Falluja, Baghdad, with the 8th Marine regiment in 2004, it was Hespan — a colloid solution — that saved lives. During a sustained battle, Jadick put the far-forward concept to its extreme, setting up an aid station in an abandoned prayer room, with stretchers up on cement blocks and bullets ricocheting outside. Colloids such as Hespan contain starch molecules that pull water into the veins from other tissues. "Hespan was a huge player," says Jadick. "It keeps fluid in the vein. But you are going to see a huge increase in combat survival when you can get blood or something like it on the battlefield."

Surgeons such as Jadick know that the most severely injured casualties — about 10% — are not only losing blood but fighting shock, hypothermia and, unhelpfully, lowered blood clotting. The answer is to get the blood volume back up — but with how much of what fluid on what timeline? Jadick had only Hespan to hand, but medics at units farther back from the action have access to various blood fractions and crystalloids — solutions of water and electrolytes — in addition to colloids.

Pressure points

Civilian doctors generally use crystalloids and colloids initially and transfuse blood or plasma later. But as of 3 January, military policy has been to give the patients something close to real blood as quickly as possible. This might mean warm blood donated by a fellow soldier or an infusion of thawed red blood cells and plasma (in a one-to-one ratio). According to Holcomb, Hespan and other products get the blood pressure back up, but do nothing to restore clotting. The clotting factors in human blood or blood products are what is needed. And because keeping blood on ice is tricky on the battlefield, other technologies will be needed.

Cellphire — a company based in Rockville, Maryland — is freeze-drying platelets, which help with clotting, to lengthen their storage time. At room temperature, platelets keep for

J. HOLCOMB

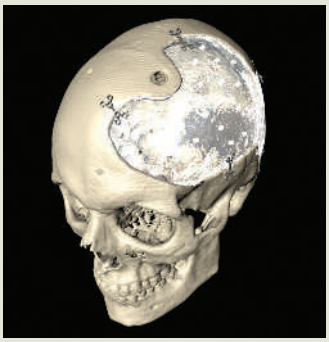
After the battle

Five or six men with high and tight military haircuts are working out in the physical therapy room of Walter Reed Army Medical Center in Washington DC. They are manipulating large balls, hopping, working the Stairmaster — and they've got maybe three legs between them. Today they are being inundated with visitors as Walter Reed tries to solicit better publicity after its outpatient care became the target of media and political outrage.

But even though their housing unit is mouldy and harbours vermin, and the hospital is leaving Washington DC in 2011, work is continuing on a new amputee ward. Iraqi veteran amputees have received so much attention from the US media that it comes as a surprise to learn that as of February, there were only 562 of them. Still, advances in prosthetics keep coming, including prosthetic arms that operate from signals sent from the nerves that once controlled the missing limb, and have been transplanted to the upper chest.

The most common wounds are those caused by explosive devices. Estimates suggest that more than 5,000 of the US wounded could have brain injuries, some of whom will benefit from a new tool called 'rapid prototyping'. This technique is used by engineers to make quick three-dimensional prototypes from computer models. At Walter Reed, the inputs are computed tomography (CT) scans of injured heads and hips. Instructions are sent to a machine that builds a model by laying down thousands of 0.125-mm layers of plastic.

Several floors below the physical therapy room, Stephen Rouse



Models are used to guide surgery.

offers visitors candy out of a model of his own cranium before explaining how the prototypes can help. If surgeons know exactly what to expect before they open up a patient, they can reduce operation times and improve outcomes. Models are also used to construct customized implants. Rouse hopes soon to buy a machine that will make titanium implants directly from a CT scan, so they will be a perfect fit.

Rouse says Walter Reed has performed several hundred craniectomies on soldiers, all of whom would have died in previous conflicts. He also stresses the positive psychological effects a skull implant has on these patients, who suffer from personality

S. ROUSE

only a few days, but the freeze-dried products can be rehydrated with sterile solution when the need arises. Cellphire hopes to have their product in clinical tests in 18 months.

Joint effort

Holcomb stresses that no advance can save lives on its own. Speed won't help, if you don't have the right devices. All the blood in the world is no good if your body armour has let you down. The goal of military medicine is moving the right people to the right places at the right time — what Holcomb calls "the trauma systems approach".

The Joint Theater Trauma Registry should reveal how well all the parts of this system work. Body protection, for example, has improved enormously since the flak jackets of Vietnam. Materials science has created ever-better protective gear — such as lighter Kevlar vests that are more resistant to bullets and shrapnel. Ceramic inserts and add-ons are available for the upper arms, neck, groin and other areas. As a result, wounds now tend to be to the arms, legs and head, with far fewer chest injuries.

The amount that body armour has contributed to the lower case fatality rate in Iraq and Afghanistan is unknown, as are many other figures that might reveal the truth behind the lower case fatality rate. Despite the lack of data, Holcomb and others argue that battlefield medicine has consistently produced medical advances, some of which transfer to civilian practice.

Although US civilians might have wars to thank for helicopter pads on hospitals and procedures for massive trauma, Iraqi citizens are getting the trauma without the advances. US care for wounded Iraqis is often no more than a goodwill gesture. General healthcare is bad. Worse, the poor security situation is leading Iraqi doctors to flee the country: up to 30% of qualified physicians are estimated to have left since 2003. Many doctors are targeted for threats and attacks by unknown parties and for unknown reasons, and a depressingly high number have been murdered or kidnapped.

J. MOORE/AP

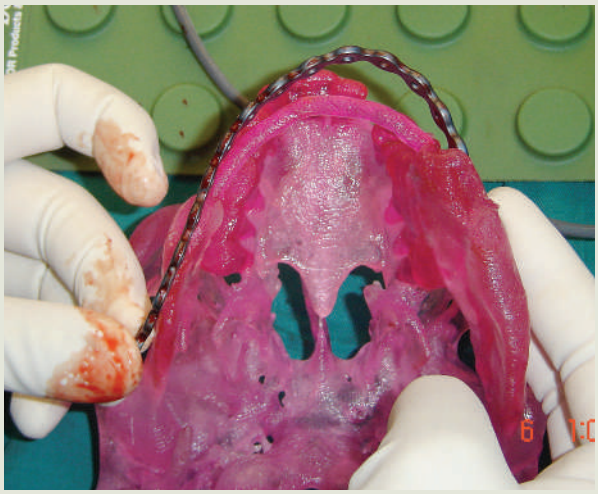
Combat support hospitals occasionally provide care for Iraqi civilians.

All wounds leave scars

Bassim Irheim Mohammed Al Sheibani of the Diwaniyah College of Medicine in Iraq wrote a letter to the *British Medical Journal* last October, explaining that there are no drugs, no trained people and no equipment[3]. No aid or assistance has been forthcoming from abroad, apart from some donations from the occupying force, which Sheibani says are often inappropriate. As for personal security, there isn't any. Sheibani himself has received phone threats. His college is planning to provide training courses in emergency medicine for students and doctors, but they lack funds and support.

As for the US military, people are coming home who would have died in previous wars, but some of them are in pretty bad shape (see 'After the battle'). For one doctor, at least, this can seem a mixed blessing. "People do fixate on the body count," says Jadick, "and there are a lot of people who are coming back alive and — I don't know, they maybe shouldn't be. You've always got to think — did we do them any favours? There are some of them that I'm not sure." ■

Emma Marris is a reporter for *Nature* based in Washington DC.



O. AL ANI

This model was used to aid surgery on an Iraqi civilian.

changes and depression. And he is convinced that the effort on the military side will transfer to civilian practice, noting that thousands of 15–20-year-olds die every year through trauma.

A surprising number of Iraqi civilians have benefited from rapid prototyping technology, thanks to a group calling itself Rapid Prototyping For Baghdad in partnership with Médecins Sans Frontières and an Iraqi dentist. Some 70–90 patients have been treated in Iraq and because of worsening security another 20–30 have been flown to Amman, Jordan, for surgery.

Despite these successes, patients are not always willing to come forward. "Patients are afraid to go to the hospitals," says Omar Al Ani, the dentist. "Before, there was crowding and a very big waiting list for the operations. The bombing and blast injuries have increased, but they are not coming because they are afraid." **E.M.**

1. Holcomb, J., Stansbury, L. G., Champion, H. R., Wade, C. & Belamy, R. F. *J. Trauma* **60,** 397–401 (2006).
2. Pruitt, B. A. Jr *Ann. Surg.* **243,** 715–729 (2006).
3. Al Sheibani, B. I. M., Hadi, N. R. & Hasoon, T. *Br. Med. J.* **333,** 847 (2006).

Poverty reduction must not exacerbate climate change

SIR — Current international policy responses to the world's two most serious problems, poverty and climate change, seem to be pulling in opposite directions.

To boost economic growth in less developed countries, international development agencies and the World Trade Organization at the Doha conference in 2001 called for a reduction in European Union and US food subsidies and the removal of trade barriers. This drive towards globalization, resulting in a free flow of trade across borders, should benefit the poorest nations through increased access to global export markets, investment capital and improved technology. These in turn should increase job opportunities and reduce poverty. According to the World Bank's 2003 report, international tourism is expected to play a major role in sustainable development of all the 49 least developed countries, where it currently provides the primary source of foreign exchange earnings.

At the same time, the pressing need to reduce greenhouse-gas emissions is becoming ever clearer. Politicians can no longer evade a range of tough policy choices, as outlined in your Editorial "Light at the end of the tunnel" (*Nature* **445,** 567; 2007).

International scientific and political debates on mitigating climate change through reducing emissions have begun to focus on the transport sector. This accounts for some 20% of global CO_2 emissions, a share that is rising rapidly. In response, political pressure by environmental groups and concerned consumers is pushing retailers to reduce 'food miles' — the distance that products travel from field to plate — by sourcing food locally. Tourism is also being targeted by environmental campaigners, as a major contributor to climate change through aircraft emissions. International aviation and shipping are currently excluded from the national emission-control targets set under the Kyoto Protocol. However, it seems increasingly unlikely that such an exclusion from future agreements will be tenable, in the face of strong evidence for the seriousness of potential climate change.

A major disjunction is, therefore, looming between climate policy and strategies for reducing poverty. This illustrates the scale of the tough choices now facing politicians. Attempts to control greenhouse-gas emissions may restrict international trade and tourism, removing the key strategies for less developed countries to grow their way out of poverty. Ironically, the same countries are most vulnerable to the direct impacts of climate change and have the lowest capacity to adapt.

Climate change therefore seems to demand that the international community should find new ways of transferring wealth from rich to poor countries, instead of relying on ever-increasing volumes of trade and tourism with their associated greenhouse-gas emissions.

Terence P. Dawson, Simon J. Allen
Centre for the Study of Environmental Change and Sustainability, University of Edinburgh, West Mains Road, Edinburgh EH9 3JN, Scotland, UK

Genetic test may lead to waste of healthy embryos

SIR — Your News Feature "The first cut" (*Nature* **445,** 479–480; 2007) highlights a controversy about preimplantation genetic diagnosis (PGD), a procedure to test for genetic disease that is now more widely used for screening chromosomes in embryos after *in vitro* fertilization (IVF). Randomized clinical trials are needed to determine whether aneuploidy screening improves pregnancy rates, or whether embryo biopsy has any impact on the health of future children and adults. There is less doubt, however, that PGD encourages the waste of healthy embryos. Your News Feature pointed out that this can potentially happen if mosaic embryos normalize in culture, but a significant fraction may occur from misdiagnosis.

Chromosomes are identified by spots of colour from fluorescent *in situ* hybridization (FISH), which has 90–92% accuracy, by some estimates. Is this good enough? At present, about half the karyotype is usually screened: nine autosomes plus two sex chromosomes. But, for every additional chromosome examined, the probability of a false result increases until, when all are screened, more than half of all embryos might be rejected, even if the accuracy is as high as 96% (0.96^{24}). If this prospect is not discouraging enough, then consider the FISH data themselves, even from the best screening centres.

During cell division, chromosome pairs normally segregate equally at opposite poles. When separation is premature or fails to happen, we expect to find an equal number of cells with an extra or a missing chromosome (monosomy), which is borne out by studies of human oocytes (see F. Pellestor *et al. Hum. Genet.* **112,** 195–203; 2003). However, FISH data consistently show a significant bias and a two- to threefold excess of monosomy compared to trisomy. A higher error rate is anticipated for monosomy as hybridization can fail for many reasons, whereas an artefactual third signal is less likely. More perplexingly, however, the incidence of aneuploidy for each chromosome is so high that one wonders whether any embryo will be judged normal when the entire karyotype can be screened by this method.

Until a more reliable technology emerges, it is likely that many embryos will be needlessly discarded. This is perverse, as the goal is to improve pregnancy rates by selecting the best-quality embryos for transfer, particularly for older patients. The high incidence of aneuploidy in our species has encouraged this application of PGD, but best interests may not be served by liberal application and literal interpretation of current FISH technology. Sometimes a patient has no diploid embryos, causing her IVF cycle to be abandoned. Perhaps she should have any embryos with monosomy transferred, unless they are XO. At worst they will be nonviable, but at best they may produce a healthy child.

Roger Gosden
Weill Medical College, Cornell University, New York, New York 10021, USA

Intellectual edge can be gained in translation

SIR — The Naturejobs Special Report "Lost in translation" (*Nature* **445,** 454–455; 2007) discussed the importance of good English to a scientific career, as well as the difficulties experienced by scientists whose first language is not English. Having worked in four countries, and having a lab of eight people from seven different countries, I have found that languages are living entities that evolve continuously.

English is under the constant influence of other languages and cultures: 'latte', for example, now appears in the Oxford English Dictionary as well in cafés around the United Kingdom. Ironically, non-native speakers may communicate more easily in English with each other than with people whose only language is English. The key to communication is to learn languages: Latin, to start with, as it is a living component of scientific vocabulary. While we are at it, we may as well pronounce Latin as its native speakers did, not with an English accent. Any language will do, though, because speaking more than one is a great mental exercise. Relaxing into the comfort zone of one's native language, on the other hand, is a step towards mental lethargy.

We must all make an effort to read, write and speak good English in order to improve our communication in science. And I make no excuses for scientists who, like myself, come from countries where English is not the first language. At the same time, we should increase awareness of diversity. Cultural diversity, like genetic diversity, is one of our most precious assets. We must cherish, nurture and preserve both of them on a global scale. One way is to improve education, thus promoting tolerance.

Francesco Colucci
The Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT, UK

Readers are welcome to comment at http://blogs.nature.com/nautilus/2007/02/languages_of_science_1.html.

COMMENTARY

DORLING KINDERSLEY/GETTY

Homeopathy has changed little since the early nineteenth century.

Science degrees without the science

Some UK universities offer science degrees in complementary medicine. **David Colquhoun** argues that these are not science but anti-science, and asks who is to blame.

Both of the exam questions shown in the box below sound pretty archaic. The first was set by Sydney Ringer in the materia medica and therapeutics exam at University College London in 1863. But what about the second? Was that too set in a medical exam in the early nineteenth century, at a time when cholera was common in London and was supposed to be spread by evil-smelling vapours (miasmas)? No. It was set in 2005 by the University of Westminster, London, in the exam for homeopathic materia medica 2A.

The least that one can expect of a bachelor of science (BSc) honours degree is that the subject of the degree is science. Yet in December 2006 the UK Universities and Colleges Admissions Service advertised 61 courses for complementary medicine, of which 45 are BSc honours degrees. Most complementary and alternative medicine (CAM) is not science because the vast majority of it is not based on empirical evidence. Homeopathy, for example, has barely changed since the beginning of the nineteenth century. It is much more like religion than science. Worse still, many of the doctrines of CAM, and quite a lot of its practitioners, are openly anti-science. For example, Christine Barry, a medical anthropologist at Brunel University, is not unusual in questioning "the supposed objectivity of scientific, biomedical forms of evidence" when writing a paper[1,2] on 'The role of evidence in alternative medicine'.

A typical exposition of the 'principles' of acupuncture sounds like science but isn't: "There are 14 major avenues of energy flowing through the body. These are known as meridians. The energy that moves through the meridians is called Qi. Think of Qi as 'The Force'. It is the energy that makes a clear distinction between life and death. Acupuncture needles are gently placed through the skin along various key points along the meridians. This helps rebalance the Qi so the body systems work harmoniously." I suppose, to the uneducated, the language sounds a bit like that of physics. But it is not. Some of the words are borrowed from science but they are used in a way that has no discernible scientific meaning whatsoever.

Yet this sort of gobbledygook is being taught in some UK universities as though it were science. The 45 'BSc' honours degrees in CAM come from 16 universities. The worst offender is the University of Westminster, with 14 BSc CAM courses. Homeopathy is the most obvious delusion because the 'medicine' contains no medicine[3]: that was obvious to Oliver Wendell Holmes[4] in 1842. Yet five of the 45 BSc degrees are offered in homeopathy. These come from the Universities of Westminster, Central Lancashire and Salford.

Can you date these exam questions?

1. *Describe the preparations of Sulphur used for cure of the itch insect. Give the different methods of applying them.*

2. *Psorinum and Sulphur are Psoric remedies. Discuss the ways in which the symptoms of these remedies reflect their miasmatic nature.*

No medicine

Other CAM courses are in aromatherapy, acupuncture, traditional Chinese medicine, herbal medicine, reflexology, osteopathy, therapeutic bodywork, naturopathy, Ayurveda, shiatsu and qigong. None of these is, by any stretch of the imagination, science, yet they form part of BSc degrees. They are not being taught as part of cultural history, or as odd sociological phenomena, but as science. The University of Westminster also offers a 'BSc' in nutritional therapy. Proponents of 'nutritional therapy' have been known to claim that changes in diet can cure anything from cancer to AIDS. For example, the British nutritionist Patrick Holford infamously recommends vitamin C as a remedy for HIV and AIDS[5,6]. Yet, Holford's Institute of Optimum Nutrition in London

offers 'top-up modules' for a BSc accredited by the University of Bedfordshire. Similarly, the Centre for Nutrition Education and Lifestyle Management offers a BSc in Nutritional Therapy 'validated' by Middlesex University. University validation of private institutions clearly doesn't work.

In 1992, John Major's government gave the title of 'university' to what had previously been polytechnics or colleges of higher education. These institutions were largely devoted to teaching. They did little research so were quite different from what had been called a university before then. They also brought with them some courses that were not remotely like those taught in older universities. This has given rise to talk of 'mickey mouse' degrees in, say, golf-course management, baking or embroidery that do not have the intellectual content of degrees in maths or French. Their existence shows that universities have changed a lot. One can argue about whether that is a good thing, but that does not matter much for the present argument.

UNIV. WESTMINSTER

Some universities, including the University of Westminster, award science degrees in homeopathy.

Teaching anti-science

What matters here is that degrees in things such as golf-course management are honest. They do what it says on the label. That is quite different from awarding BSc degrees in subjects that are not science at all, but are positively anti-science. In my view, they are plain dishonest. One sad consequence of this is the enormous harm it does to the reputations of the 16 post-1992 universities that run CAM degrees in their very proper wish to be recognized as comparable institutions to the older universities. A few older universities also have departments that teach anti-science as fact (for example, complementary medicine groups at the Universities of Southampton and York and the Open University CAM course) but at least they don't award degrees in the subject.

Why don't regulators prevent BSc degrees in anti-science? The Quality Assurance Agency for Higher Education (QAA) claims that "We safeguard and help to improve the academic standards and quality of higher education in the UK." It costs taxpayers £11.5 million (US$22 million) annually. It is, of course, not unreasonable that governments should ask whether universities are doing a good job. But why has the QAA not noticed that some universities are awarding BSc degrees in subjects that are not, actually, science? The QAA report on the University of Westminster courses[7] awards a perfect score for 'curriculum design, content and organization,' despite this content consisting largely of what I consider to be early-nineteenth-century myths, not science. It happens because the QAA judges courses only against the aims set by those who run the QAA, and if their aims are to propagate magic as science, that's fine.

"It seems that we can't expect anything useful from regulators."

The chief executive of the QAA, Peter Williams, is quite right when he says that one would not want a government agency dictating to universities what degrees they should or should not provide. But in that case, the QAA seems to be totally redundant. Their own rules prevent them from doing anything useful. Today, universities are beset by all manner of regulators, assessors and rankings. They cost a great deal of time and money, for little noticeable gain. All of us who do research (rather than talk about it) know the disastrous effects that the Research Assessment Exercise has had on research in the United Kingdom: short-termism, intellectual shallowness, guest authorships and even dishonesty. It seems that we can't expect anything useful from regulators. The real responsibility lies with university vice-chancellors, and it is they who must take the blame.

Who's in charge?

Why do vice-chancellors allow their universities to award BSc degrees in anti-science? In 2003 I wrote twice to the vice-chancellor of Westminster about the teaching of such degrees, but got no response. One reason is presumably that degrees in anti-scientific subjects exist because there is a public demand for them. They are vocational degrees and people are smart enough to know that magic is a good way to make money. The universities make money too, by getting bums on seats. So what's the problem? Some vice-chancellors seem to have lost sight of what universities are for. Before 1826, universities in England, but not Scotland, were largely theological seminaries; after the foundation of the University of London in 1826 (later University College London), universities became places where people sought, as best they could, to discover the truth[8]. They became places you could turn to for independent thought and opinions, undistorted by financial interests. The best ones still are, but that independence of thought has never been more at risk. The pressure now is not towards theology, but towards corporatization. If newspaper editors can make money from astrology columns, why shouldn't universities get in on the act? A few of them have, but in doing so they become the antithesis of what a university should be.

The Committee of Vice-Chancellors and Principals (renamed 'Universities UK' post-1992) has been remarkably unhelpful. Their only intellectual contribution to the debate appears to be in a 2003 report on Universities and the NHS[9] in which Martin Harris, then vice-chancellor of Manchester University, says "Until very recently, so-called 'alternative therapies' have been viewed with scepticism by the medical profession. However, more doctors and nurses are now recognizing that these treatments can and do work."

One wonders who gave him that idea. An astrologer perhaps? Or a homeopath? ■

David Colquhoun is in the Department of Pharmacology at University College London, UK.

1. Colquhoun, D. *DC's Improbable Science Page* www.ucl.ac.uk/Pharmacology/dc-bits/quack.html.
2. Barry, C. A. *Soc. Sci. Med.* **62,** 2646–2657 (2006).
3. Editorial *Lancet* **336,** 690 (2005).
4. Holmes, O. W. *Homeopathy and its Kindred Delusions* www.quackwatch.org/01QuackeryRelatedTopics/holmes.html (1842).
5. Goldacre, B. *The Guardian* 17 February (2007).
6. www.patrickholford.com/content.asp?id_Content=1778
7. QAA *Other Subjects Allied to Medicine* www.qaa.ac.uk/reviews/reports/subjectlevel/q171_01.pdf (2000).
8. Desmond, A. *The Politics of Evolution: Morphology, Medicine and Reform in Radical London* (Univ. Chicago Press, 1992).
9. Harris, M. *Universities and the NHS: Partners in Care* http://bookshop.universitiesuk.ac.uk/downloads/partnersincare.pdf (2003).

Acknowledgement. I am grateful to P. Williams of the QAA for information and help.

BOOKS & ARTS

The interior designer

Can the physiological agents of homeostasis create the appearance of design in nature?

The Tinkerer's Accomplice: How Design Emerges From Life Itself
by J. Scott Turner
Harvard University Press: 2007. 304 pp.
£14.49, $27.95

Claus Wedekind

Sharing a broadly accepted idea or philosophical concept comes with a danger: after a period of indulgence in mutual affirmation, it is easy to forget how to effectively defend the concept against a smart and captious critic. Established politicians sometimes stumble and get lost in clumsy arguments when forced to defend the basic concepts of their politics against a cleverly presented and maybe radically different opinion. And evolutionary biologists can struggle to find their best arguments when challenged by a well-prepared enthusiast of 'intelligent design'. Non-physiologists, for example, might overlook the agents of homeostasis that lead, largely by themselves, to the marvellous harmony of structure and function we observe in nature.

In his book *The Tinkerer's Accomplice*, Scott Turner provocatively calls this harmony of structure and function 'designedness', probably because, as he writes, there is "no better way to open minds than to irritate them a bit". And he does an excellent job here, not just with the irritation part but also with what follows.

Most chapters start with a pictorial description of a rather randomly picked phenomenon to build up to a question that may not have been obvious to the average reader before but becomes really burning now. Turner then leads us competently and in great detail through a likely physiological explanation. The argument is well organized, with a high density of information, but the accessible and often humorous language carries us along. When he explains, for example, the thoughts and observations that led to his understanding of how termite chimneys not only capture wind to power ventilation but also regulate its capture, and how this makes the chimney an organ of homeostasis, his narrative skills make you feel as if you were sitting with him at a fireplace in the African bush. When he explains the self-organization of blood vessels, or the way the digestive tract develops and continuously adapts its functionality, you might feel as if you are in the office of a senior professor who is carried away by his enthusiasm for the subject. And you might start pondering how it would feel to grow an antler after reading that its developing shape interacts with the nervous system in such a way that experimentally induced modifications of the antler's shape are memorized in the brain and influence the growth of next year's antlers, even after the modified antler is cast.

Can homeostasis explain the design of termite chimneys?

O. NEWMAN/NATUREPL.COM

Turner challenges Theodosius Dobzhansky's notion that "nothing in biology makes sense except in the light of evolution" with a claim that could be summarized as: "no attribute of life, including its evolution, really makes sense unless we view it through a physiological lens". He reintroduces various dynamic and self-organizing systems and explains the concept of 'Bernard machines', named after Darwin's contemporary Claude Bernard, who emphasized the role of homeostasis in physiology. Bernard machines create environments and regulate them, just like those termite chimneys. For example, fibroblasts remodel collagen meshes to conform to imposed loads, and embryonic cells form into sheets that fold and create new physiological environments.

This leads to the tantalizing question of whether darwinian evolution can dismiss intentionality. Obviously, creative brains can cope better with an unpredictable world and may have a selective advantage, so creativity and intentionality can evolve and in turn influence evolution. But does it really need a brain like ours to bring intentionality into play? Turner views this question through a physiological lens and develops a picture of a modular brain that could be understood as a kind of 'climax' ecosystem with competing and coevolving cells, and with homeostasis as the organizing principle of cognition. He argues that we intentionally design the world when our neural ecosystems generate ideas that then guide our bodies to reshape it. The point is that the brain may be just one example of what Turner calls 'persistors' — persistent environments that are created by systems of Bernard machines and that have a process-based form of heritable memory. 'Darwin machines' — replicators that have to prove themselves under natural selection — shape evolution in the absence of intentionality. But the author argues that life and evolution happen when Darwin machines act in concert with Bernard machines, which are the agents of homeostasis and can be seen, in their own particular way, as goal-seeking and purposeful. These are the 'tinkerer's accomplices' of the title.

It is fun to read Turner's prose, to learn from him about self-organizing systems and their enormous significance in evolution, and to think through his arguments, with all their accompanying intellectual challenges. This important book is for those who search for an understanding of the various forms that life can take and of how life works. It is also a wonderful book for physiology students, especially for those who could use a motivational kick to help them continue their studies. ■

Claus Wedekind is in the Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland.

The age of chance

Uncertainty: Einstein, Heisenberg, Bohr, and the Struggle for the Soul of Science
by David Lindley
Doubleday: 2007. 272 pp. $26

Arthur Fine

In *Uncertainty*, David Lindley tells the intriguing tale of how Albert Einstein, Werner Heisenberg and Niels Bohr (among others) struggled to create and understand the new quantum physics. Lindley organizes his tale around the issue of indeterminism, which Max Born raised in 1926 in the paper that introduced probability as fundamental to interpreting the quantum world. Within a year, at the end of his paper on the uncertainty principle, Heisenberg declared determinism (or causality) dead, a pronouncement that brought probability, chance and uncertainty into the quantum domain in a fundamental way.

Lindley tracks the rise of chance from its roots in statistical reasoning (brownian motion and entropy) through to Bohr's 'jumping planetary model' of the atom and beyond. He selects important episodes from this 'old' quantum theory and then retells them in a lively and insightful manner. This provides the background for Heisenberg's theory of matrix mechanics and Erwin Schrödinger's wave mechanics. The author tells how Bohr encouraged, derided, cajoled, inspired and browbeat all sides to orchestrate the Copenhagen synthesis to meet his own physical intuitions and philosophical likings. Lindley captures the passion of the struggle, showing both the public controversies and the sometimes harsh private judgements (for example, writing to third parties, Heisenberg and Schrödinger each described the other's work as repulsive, and worse).

What elicited this passion is the question of whether the behaviour of atoms (quanta) can be described continuously in space and time. Heisenberg's strange matrix methods are driven by his belief that they cannot, whereas the wave approach opens up the possibility that they can. That prospect, and the divisive reaction to wave mechanics, was bolstered by a development that Lindley scarcely mentions. It concerns Louis de Broglie.

Lindley does mention de Broglie, but describes his brilliant association of waves with particles as merely an "elementary idea that crossed his mind". This is a stab-in-the-dark history that rips the important de Broglie relations from their theoretical context. That context was de Broglie's proposal for a radically non-newtonian dynamic theory in which particles followed trajectories determined by an associated wave. De Broglie showed how this could account for basic interference phenomena (as waves) as well as quantized energy (as particles). That drew it to the attention of Einstein and Schrödinger, who found the ideas promising, and to Heisenberg, Pauli and Bohr, who felt threatened by them. But de Broglie lacked a general treatment of the waves that guided his particles, and that is where wave mechanics comes in.

The fifth Solvay conference, held in Brussels in October 1927, was a major event involving nearly all the figures mentioned in Lindley's story. Lindley describes the struggle between updated versions of wave and matrix mechanics, and Einstein's informal challenges. The confrontation also included an updated de Broglie theory, with continuous space-time trajectories. Moreover, full determinism was guaranteed by linking de Broglie's new dynamics with Schrödinger's waves. Lindley's book misses all this, misrepresenting de Broglie as fighting a rearguard action on behalf of Schrödinger. It's as if the head-on collision over determinism, the introduction of a radically different dynamics, the discussions, thought experiments, probing challenges and responses never occurred.

Lindley touches on several significant later events in his story, including an encounter between Einstein and Bohr during the 1930 Solvay conference, reactions to the Einstein–Podolsky–Rosen paradox and an update on Schrödinger's cat. Other sections also mention a few sociological and philosophical reactions, which Lindley criticizes and quickly dismisses, along with any dissent from the orthodox view. These sections are disappointing. It seems, unfortunately, that Lindley's perceptive and sympathetic treatment of ideas and figures is reserved for the winning side of the struggle. He marginalizes Einstein's concerns by presenting him as a young revolutionary turned old reactionary. Lindley keeps returning to Einstein's desire for causality, while downplaying Einstein's equally strong insistence that physics, causal or not, should deal with nature itself, not just our observations. The author also overlooks Einstein's radical critique of the classical, physical concepts used in quantum theory, as well as his programme for developing new concepts, and his eventual openness to an algebraic, rather than a spatiotemporal, setting for the physics (not exactly a reactionary agenda).

For a different perspective, readers may be interested in David Cassidy's *Uncertainty: The Life and Science of Werner Heisenberg* (W. H. Freeman, 1991), Mara Beller's *Quantum Dialogue* (University of Chicago Press, 1999), and two books by Abraham Pais, *Subtle is the Lord* (Oxford University Press, 1982) and *Niels Bohr's Times* (Clarendon, 1991). A forthcoming book by Guido Bacciagaluppi and Antony Valentini, *Quantum Theory at the Crossroads* (Cambridge University Press, 2007; available online at http://arxiv.org/abs/quant-ph/0609184), will offer a reanalysis of the 1927 Solvay conference.

Uncertainty tells the tale of the struggle over quantum theory from the perspective of an omniscient narrator. The narrator turns out not to be all that impartial. He is also not omniscient, stumbling a few times (and not just by omission), as anyone might in such a complex tale. Yet the story is told with verve, has some interesting historical asides, and makes a good read. Still, you might not want to believe it; well, certainly, not all of it. ■

Arthur Fine is in the Department of Philosophy, University of Washington, Seattle, Washington 98195-3350, USA.

P. EHRENFEST/AIP EMILIO SEGRE VISUAL ARCHIVES

Niels Bohr (left) and Albert Einstein sought to shape the development of quantum physics in 1930.


D. PEARSON/ALAMY

Indecent exposure: the widespread use of toxic chemicals puts people's health at risk.

Poisoned at work

How Everyday Products Make People Sick: Toxins at Home and in the Workplace
by Paul D. Blanc
University of California Press: 2007. 385 pp. $50, £32.50 (hbk); $19.95, £12.95 (pbk)

Anthony Robbins

The title of Paul Blanc's book *How Everyday Products Make People Sick* makes it sound like a consumer or self-help guide, and gives no hint of its literate and charming contents. It is focused on the history of workplace-caused illnesses and how, from the earliest times, they presaged community epidemics. Using colourful stories, Blanc offers evidence for his main points, most notably that there is no absolute division between consumers and workers. It is a wonderful read.

Blanc, a specialist in occupational and environmental medicine, treats "diseases that people get from their work or as a result of pollution". This medical specialty is distinguished from others not by the signs and symptoms it observes, for the human body has a limited number of ways of reacting to poisons and pathogens, but by the histories describing the patient's exposure: what, when, for how long, in what concentration, and by what route? Treatment includes variations on normal medical intervention, often extended to strategies for preventing further exposure.

Blanc builds his arguments around historical observations, suggesting that we should not be surprised when exposure reveals some new product or chemical to be poisonous. A hint from the past should alert us to future problems. In the first chapter, he relates some "forgotten histories of 'modern' hazards" about mercury, air and water pollution, asbestos, job burnout and the increasing frequency of carpal tunnel syndrome. Blanc never tries to be exhaustive in his exposition of a problem, but colourful instead, telling the reader about enquiring scientists and never omitting the unfortunate victims.

The workplace offers a good way to learn about toxic substances because workers in one plant or job are often distinct from everyone else in terms of their exposure to chemicals. Adverse effects first seen in industry are often seen later in homes and the environment. For example, the search to produce white cloth drove bleaching processes for centuries before chlorine bleaches caused a revolution in the textile industry globally, making chlorine ubiquitous — an engaging story the way Blanc tells it. He uses equally entertaining tales of glue, solvents and "job fever" to illustrate the important role of commerce.

Industry has driven many changes in the workplace, the home and the environment. Any ill effects of exposure to chemicals are, as a rule, detected very quickly, but society acts much more slowly to prevent or correct them. Industry introduces new processes, involving exposure to new chemicals, then routinely opposes any investment aimed solely at preventing further exposure. "In the early years of the twentieth century," writes Blanc, "medical experts shared a sense that the scourge of carbon disulfide had passed, viewing this as a public health victory that was incorrectly attributed to improved regulatory controls, when actually the reduced use of carbon disulfide in vulcanization was due to new technologies, not government action." One theme recurs: public-health regulators are always playing catch-up, be they the do-gooders of sanitary reform in England or the hamstrung US regulatory agencies of the early twenty-first century.

Near the end of the book, Blanc predicts future problems, using manganese as an example. Manganese is poisonous in the body, but as its common inorganic form is poorly absorbed by the gastrointestinal tract, damage has been limited. New organic chemicals that incorporate manganese are another story, however. Sales of two relatively new pesticides containing organic manganese, Maneb and Mancozeb, already exceed £11 million (US$21 million) annually. Blanc suggests caution and watchfulness.

Blanc brings his book up to date, noting, for example, industry-sponsored data reanalysis apparently intended to muddy the scientific picture and delay regulation. His conclusions rely heavily on the current state of public-health and regulatory affairs in the United States. A slightly broader sweep of enquiry might have also indicted Europe, Japan and developing countries, as well as the World Trade Organization and other international bodies.

In the 26 years since I directed the US National Institute for Occupational Safety and Health, I haven't seen a book that so clearly describes how the health of workers fits into the big picture, and how occupational health can also protect the public. I will recommend it to all my public-health students and to colleagues who might appreciate a way to understand the forest while they study the trees.

Blanc's emphasis on the economy triggers several questions. Might we be able to construct a dynamic picture of exposure to poisons? Using a Leontief input–output matrix model of an economy, which can predict the effect of changes in one industry on others, would it be possible to incorporate information on toxicity, number of exposures and populations exposed? As industry introduces new products, changes production processes, relocates plants, substitutes ingredients and installs protective technologies, perhaps we would be able to foresee the consequences for workers and the environment. And maybe protection will come sooner. ■

Anthony Robbins is co-editor of the *Journal of Public Health Policy*, 213 West Canton Street, Boston, Massachusetts 02116, USA.

Frontier at your fingertips

Between the nano- and micrometre scales, the collective behaviour of matter can give rise to startling emergent properties that hint at the nexus between biology and physics.

Piers Coleman

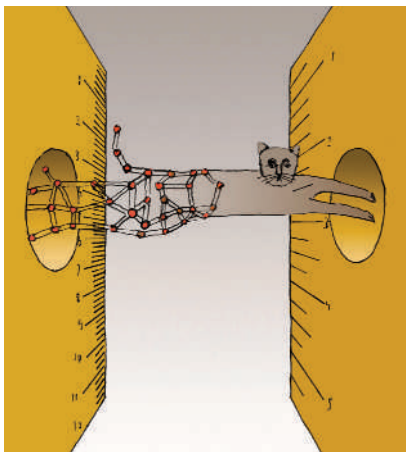
The Hitchhiker's Guide to the Galaxy famously features a supercomputer, Deep Thought, that after millions of years spent calculating "the answer to the ultimate question of life, the Universe and everything", reveals it to be 42. Douglas Adams's cruel parody of reductionism holds a certain sway in physics today. Our 42 is Schroedinger's many-body equation: a set of relations whose complexity balloons so rapidly that we cannot trace its full consequences up to macroscopic scales. All is well with this equation, provided we want to understand the workings of isolated atoms or molecules up to sizes of about a nanometre. But between the nanometre and the micrometre wonderful things start to occur that severely challenge our understanding. Physicists have borrowed the term 'emergence' from evolutionary biology to describe these phenomena, which are driven by the collective behaviour of matter.

Take, for instance, the pressure of a gas — a cooperative property of large numbers of particles that is not anticipated from the behaviour of one particle alone. Although Newton's laws of motion account for it, it wasn't until more than a century after Newton that James Clerk Maxwell developed the statistical description of atoms necessary for understanding pressure.

The potential for quantum matter to develop emergent properties is far more startling. Atoms of niobium and gold, individually similar, combine to form crystals that, kept cold, show dramatically different properties. Electrons roam free across gold crystals, forming the conducting fluid that gives gold its lustrous metallic properties. Up to about 30 nanometres, there is little difference between gold and niobium. It's beyond this point that the electrons in niobium start binding together into the coupled electrons known as 'Cooper pairs'. By the time we reach the micrometre scale, these pairs have congregated in their billions to form a single quantum state, transforming the crystal into an entirely new metallic state — that of a superconductor, which conducts without resistance, excludes magnetic fields and has the ability to levitate magnets.

Superconductivity is only the start. In assemblies of softer, organic molecules, a tenth of a micrometre is big enough for the emergence of life. Self-sustaining microbes little more than 200 nanometres in size have recently been discovered. Although we understand the principles that govern the superconductor, we have not yet grasped those that govern the emergence of life on roughly the same spatial scale.

In fact, we are quite some distance from this goal, but it is recognized as the far edge of a frontier that will link biology and physics. Condensed-matter physicists have taken another cue from evolution, and believe that a key to understanding more complex forms of collective behaviour in matter lies in competition not between species, but between different forms of order. For example, high-temperature



J. KAPUSTA/IMAGES.COM

superconductors — materials that develop superconductivity at liquid-nitrogen temperatures — form in the presence of a competition between insulating magnetic behaviour and conducting metallic behaviour. Multi-ferroic materials, which couple magnetic with electric polarization, are found to develop when magnetism competes with lattice-distorting instabilities.

A related idea is 'criticality' — the concept that the root of new order lies at the point of instability between one phase and another. So, at a critical point, the noisy fluctuations of the emergent order engulf a material, transforming it into a state of matter that, like a Jackson Pollock painting, is correlated and self-similar on all scales. Classical critical points are driven by thermal noise, but today we are particularly interested in 'quantum phase transitions' involving quantum noise: jigglings that result from Heisenberg's uncertainty principle. Unlike its thermal counterpart, quantum noise leads to diverging correlations that spread out not just in space, but also in time. Even though quantum phase transitions occur at absolute zero, we're finding that critical quantum fluctuations have a profound effect at finite temperatures.

For example, 'quantum critical metals' develop a strange, almost linear temperature dependence and a marked predisposition towards developing superconductivity. The space-time aspect of quantum phase transitions gives them a cosmological flavour and there do seem to be many links, physical and mathematical, with current interests in string theory and cosmology. Another fascinating thread here is that like life, these inanimate transformations involve the growth of processes that are correlated and self-sustaining in time.

Some believe that emergence implies an abandonment of reductionism in favour of a more hierarchical structure of science, with disconnected principles developing at each level. Perhaps. But in almost every branch of physics, from string theory to condensed-matter physics, we find examples of collective, emergent behaviour that share common principles. For example, the mechanism that causes a superconductor to weaken and expel magnetic fields from its interior is also responsible for the weak nuclear force — which plays a central role in making the Sun shine. Superconductors exposed general principles that were used to account for the weak nuclear force.

To me, this suggests that emergence does not spell the end for reductionism, but rather indicates that it be realigned to embrace collective behaviour as an integral part of our Universe. As we unravel nature by breaking it into its basic components, avoiding the problem of '42' means we also need to seek the principles that govern collective behaviour. Those include statistical mechanics and the laws of evolution, certainly, but the new reductionism that we need to make the leap into the realm between nano and micro will surely demand a new set of principles linking these two extremes. ■

Piers Coleman is in the Department of Physics and Astronomy, Rutgers University, 136 Frelinghuysen Road, Piscataway, New Jersey 08854-8019, USA.

FURTHER READING
Anderson, P. W. *Science* **177,** 393 (1972).
Laughlin, R. B. *A Different Universe* (Basic Books, 2005).
Davis, J. C. http://musicofthequantum.rutgers.edu (2005).
Coleman P. & Schofield, A. J. *Nature* **433,** 226–229 (2005).

For other essays in this series, see http://nature.com/nature/focus/arts/connections/index.html

NEWS & VIEWS

MECHANOCHEMISTRY

A reaction to stress

Brad M. Rosen and Virgil Percec

Chemists usually kick-start reactions with heat, light or electricity, but a far less common option is to use mechanical stress. It now seems that stress not only triggers reactions, but can also direct their course.

When frustrated by a failed reaction, many disheartened chemists wish that they could simply reach into their flasks and pull apart any uncooperative chemical bonds. Although molecule-sized tweezers are still not on the market, on page 423 of this issue Jeffrey Moore and colleagues[1] disclose a technique that may one day grant chemists their wish: a mechanical method for controlling the reactivity of molecules.

One of the ultimate goals of synthetic chemistry is to develop the ability to efficiently and selectively overcome all thermodynamic reaction barriers, so that any reactant can be converted to any desired product. Traditionally, chemists use catalysts, heat, light and sometimes electric fields to break down these barriers. More recently, they have mimicked nature's enzyme-accelerated reactions, either by building enzyme-like nanoscale reaction vessels[2,3] or by using the self-aggregation of reactants to modulate reactivity[4,5].

But mechanochemical activation — the acceleration and/or control of reactivity through mechanical deformation — is a method that has not taken hold. A few examples exist in organic chemistry[6], notably the stress-induced cleavage of polystyrene molecules to form reactive carbon radicals. Unfortunately, this process can't be applied to molecules other than polymers, and it also generates a mixture of radicals by breaking the polymer chain at random locations, rather than at a single desired location. This limits its synthetic usefulness.

Nevertheless, mechanochemistry holds great promise. Reactions can be thought of in terms of their 'landscapes' — multi-dimensional surfaces (much higher than three dimensions) that map the peaks and valleys of potential energy as molecules interact. Existing methods of initiating and accelerating reactions either redistribute reagents on the landscape by means of direct energy transfer, or push the reagents onto a completely different landscape through new chemical interactions. Mechanochemistry is the only technique that can actually alter the reaction landscape itself, through force.

With this in mind, Moore and colleagues[1] wondered what would happen if they selected a bond and literally pulled it apart. Their targets were two isomers of benzocyclobutene — molecules that can be thought of as hexagonal benzene rings with a square of hydrocarbons (the cyclobutene section) fused along one edge (Fig. 1). The isomers differed in the orientation of the chemical groups (in this case, methoxy groups, OCH_3) attached to the cyclobutene section. By modelling the isomers using a computational technique known as COGEF (constrained geometry simulates external force)[7], the authors predicted that pulling the methoxy 'handles' in opposite directions would cleave a bond in the cyclobutene, yielding the same product (E,E-*o*-quinodimethide diene; Fig. 1), regardless of the choice of isomer.

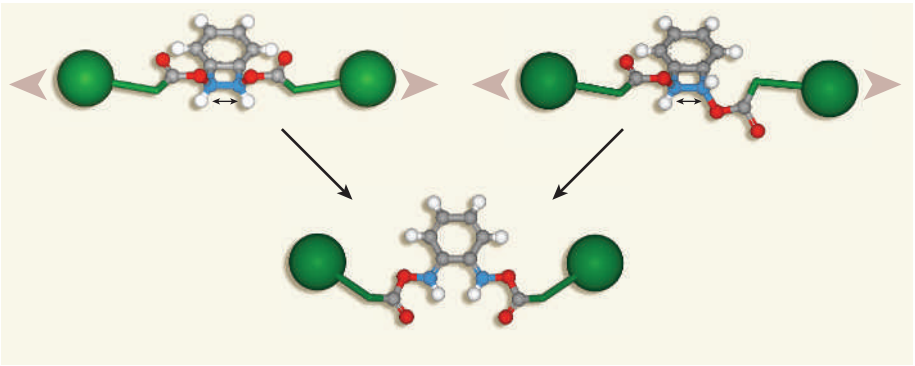


Figure 1 | Stress-induced bond cleavage. Moore and colleagues[1] show that mechanical stress (indicated by arrowheads) can be transmitted to a target bond (in blue; double-headed arrows indicate bond stretching) in benzocyclobutene molecules by means of polymer chains (green spheres) attached at either side of that bond, leading to bond cleavage. Two isomers (described as *cis* and *trans*) of the benzocyclobutene starting material are shown. According to the law of conservation of orbital symmetry, each isomer should yield a different product. In fact, each isomer gives the same product (E,E-*o*-quinodimethide diene), demonstrating that mechanical force not only activates the reaction, but also controls the reaction pathway.

This may seem straightforward, but there's a problem. A fundamental theory that is used to analyse reactions such as these — the Woodward–Hoffmann rules of orbital symmetry[8] — predicts that each isomer will form a different product. Moore and colleagues' models amount to chemical heresy. So how can they be rationalized? The answer is that mechanical stress reshapes the reaction pathway, allowing the reactants to proceed along different, formerly inaccessible routes that ultimately lead to the same product.

The authors' claims that the laws of orbital symmetry can be side-stepped by mechanochemistry are bold, but substantiated. The rapid collapse of solvent bubbles generated by ultrasound is known to induce mechanical stress along the axis of long polymer chains, enough to rupture a carbon–carbon bond. With this in mind, Moore and colleagues attached two identical polymer chains to their benzocyclobutenes; these polymers can channel ultrasound-induced stress directly onto a target bond, causing it to break. According to the authors' predictions, both of the polymer-containing benzocyclobutene isomers should be converted into the same E,E-*o*-quinodimethide diene product.

But it was difficult for Moore's team to prove directly that the reaction had occurred, using conventional spectroscopic techniques, because the polymer chains are much larger than the benzocyclobutene fragment. Signals from the reaction centre were too small to be detected. A sensitive technique was needed to verify the predictions. The authors therefore performed their reactions in the presence of a chemical trap that reacted with the product as it formed; the trap incorporated a tag that absorbs ultraviolet light. The authors'

analysis of the resulting products showed a marked increase in ultraviolet light absorption associated with the tag.

Although this confirmed that the products of the mechanochemical process were reacting with the chemical trap, it didn't provide information about the structure of those products or about the mechanism that yielded them. To address this problem, the authors labelled the trap molecules with ^{13}C atoms, which can be detected by nuclear magnetic resonance (NMR) spectroscopy. The NMR spectra of the products from both benzocyclobutene isomers showed the same single peak, suggesting that the trap reacts to form identical products regardless of which isomer is used as the starting material. This aberration strongly supports the authors' theory that mechanochemistry can alter reaction pathways and formally break the laws of orbital symmetry.

Although chemists might not immediately rush to adopt this technique, the work is a remarkable first step in using mechanical stress to break bonds in a synthetically useful organic reaction. The applications of polymers in chemistry have come a long way — from the resins that act as supports in solid-phase peptide synthesis[9] to the DNA templates that systematically bring reactants together[10]. In Moore and colleagues' study[1], the role of polymers has evolved further to that of a scaffold, not only encouraging reactions but also moulding the subsequent reaction pathway. The authors have provided a glimpse into a possible future where mechanical deformation, mediated by reactants attached to polymer chains, will allow access to previously unattainable molecules and materials, reducing the need for expensive and often environmentally unfriendly catalysts. ■

Brad M. Rosen and Virgil Percec are in the Roy and Diana Vagelos Laboratories, Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6323, USA.
e-mail: percec@sas.upenn.edu

1. Hickenboth, C. R. *et al. Nature* **446,** 423–427 (2007).
2. Kang, J. & Rebek, J. *Nature* **385,** 50–52 (1997).
3. Leung, D., Fiedler, D., Raymond, K. & Bergman, R. G. *Angew. Chem. Int. Edn* **43,** 963–966 (2004).
4. Percec, V. *et al. Nature* **391,** 161–164 (1998).
5. Percec, V., Ahn, C.-H. & Barboiu, B. *J. Am. Chem. Soc.* **119,** 12978–12979 (1997).
6. Beyer, M. K. & Clausen-Schaumann, H. *Chem. Rev.* **105,** 2921–2948 (2005).
7. Beyer, M. K. *J. Chem. Phys.* **112,** 7307–7312 (2000).
8. Woodward, R. & Hoffmann, R. *The Conservation of Orbital Symmetry* (Academic, New York, 1970).
9. Merrifield, R. B. *Angew. Chem. Int. Edn Engl.* **24,** 799–892 (1985).
10. Gartner, Z. J. & Liu, D. R. *J. Am. Chem. Soc.* **123,** 6961–6964 (2001).

ASTEROIDS

Spun in the sun

William F. Bottke

Two asteroids have been observed gradually spinning faster and faster, and the hot tip is that sunlight is the cause. If so, this could give us a handle on the dynamics and evolution of the asteroid belt in general.

"Captain, something is spinning up these asteroids, and there's no way we can stop it!" What might sound like an excerpt from a long-lost episode of *Star Trek* is, in fact, a pretty precise description of a genuine scientific mystery. On page 420 of this issue, Kaasalainen *et al.*[1] show that asteroid (1862) Apollo, a 1,400-metre-diameter near-Earth asteroid (one whose orbit intersects that of Earth), has noticeably increased its rate of rotation over 25 years of observation*. And in papers published in *Science*, Lowry *et al.*[2] and Taylor *et al.*[3] report ground-based optical and radar observations showing that (54509) 2000 PH5, a 114-metre-diameter near-Earth asteroid, is also spinning up.

So what's the cause, if it is not tractor beams or dilithium crystals? The answer, it seems, is sunlight. The measurements might be the first direct detection of a long-hypothesized phenomenon known as the Yarkovsky–O'Keefe–Radzievskii–Paddack (YORP) effect. This is a torque produced when sunlight from an asteroid's surface is reflected and re-emitted at thermal, infrared wavelengths[4,5]. According to the theory, these thermal torques can cause small asteroids to spin up or down with time, the direction and acceleration of the spin being determined by the shape and orientation of each body (Fig. 1). Given enough time, the YORP effect can even flip a body so it ends up spinning in the opposite direction.

The idea that solar radiation can affect asteroid dynamics goes back to the first of the eponymous YORP scientists, Ivan Osipovich Yarkovsky, a Polish civil engineer who worked for a Russian railway company by day, and bent his mind to scientific problems by night[6]. Shortly before his death in 1902, Yarkovsky published a pamphlet describing how infrared heat that was re-radiated away from the surface of an asteroid could provide a small thrust. In much the same manner, ices sublimating to their gaseous state off the surface of a comet create a 'rocket effect' that propels the comet strongly enough to change its orbit. Yarkovsky's effect works much more slowly, mainly because the angular momentum of a body changes more easily when comparatively massive gas and dust are thrown off than when infrared photons are radiated away. Nevertheless, it is now recognized that the thermal forces of the Yarkovsky effect can cause small bodies to drift slowly towards or away from the Sun to an extent determined by their size, the direction of their spin axis and several other parameters[7].

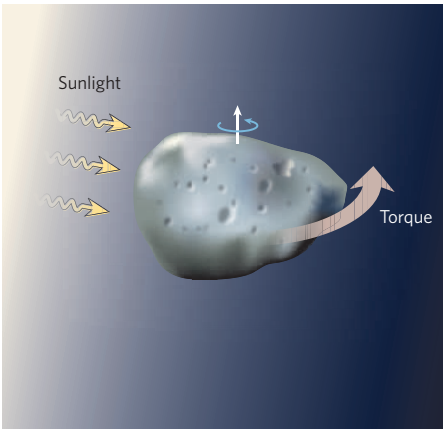


Figure 1 | YORP in action. An unevenly shaped asteroid heating up in sunlight re-radiates the energy away at right angles to its surface. The resultant net torque can both provide a small thrust (the Yarkovsky effect), causing the asteroid to drift towards or away from the Sun, and also change the asteroid's spin rate — the YORP effect, now seemingly observed for the first time[1–3].

After languishing in obscurity for several decades, Yarkovsky's work began to inspire several experiments in the 1970s, which included shining light on centimetre-sized meteoroids in a vacuum chamber[8]. The results seemed to show that sunlight reflecting off an asymmetrical object of asteroid size would not only give it an additional thrust, but also cause it to spin up over geologically short timescales. Objects spinning too fast might even undergo 'rotational bursting', shedding mass to slow themselves down.

The modern pioneer of the YORP effect is David Rubincam, who in 2000 derived the theory of how thermal torques affect real asteroids, and performed numerical simulations[4]. An important conclusion was that asteroids spin up much faster if they have some windmill-like asymmetry. In other words, propeller-shaped asteroids are more affected by YORP than are spheres or ellipsoids.

This brings us back to the contemporary YORP observations[1–3]. As irregularly shaped asteroids turn in space, they reflect different amounts of light to Earth. By carefully measuring how this light changes[1,2] — or, equivalently, how the returned beam strength from a well-aimed radar signal varies[3] — astronomers can determine the body's rotation rate and shape. Such measurements should also allow a prediction of the asteroid's orientation at any time in the future, whether it looks like a flattened

*This article and the paper concerned[1] were published online on 7 March 2007.

skipping stone (like Apollo) or a pulled tooth (like 2000 PH5). But Kaasalainen *et al.*[1] found that, 25 years after the first observations, the long axis of Apollo, around which the asteroid spins once every 3 hours, was 125° ahead of where it should have been. Similarly, Lowry *et al.*[2] and Taylor *et al.*[3] found that the 12-minute-spinner 2000 PH5 was 240° ahead of expectation after only 4 years. YORP theory not only predicts these changes, but it also explains why 2000 PH5, a smaller body more susceptible to thermal forces, was spun up faster.

So why should we care? Collisions among asteroids mean that they are already spinning in many different ways, so what does it matter if thermal radiation adds a few more twists and turns?

One answer is that the YORP effect is more efficient at changing the spin rates of irregularly shaped, kilometre-sized asteroids than are collisional impacts[5]. Indeed, it is so efficient that rotational bursting could well be an important mechanism for creating asteroid satellites both of near-Earth asteroids[9,10] and of asteroids in the main belt between Mars and Jupiter. The presence of a 75-m asteroid in Apollo's tow might well be a case in point. At the other extreme, some asteroids are spun down so far that they lose virtually all of their rotational angular momentum. This provides a firm explanation for how some asteroids enter into tumbling rotation states[5] — for example, (4179) Toutatis, which made a particularly close approach to Earth in September 2004.

A second answer is that the Yarkovsky and YORP effects work together to deliver asteroids and meteoroids to a state of so-called orbital resonance, and thence possibly to Earth[5]. When a small asteroid is created by a collision in the main belt, it immediately begins to drift inwards or outwards as a result of Yarkovsky thermal forces. The body's speed and direction, however, are controlled by the orientation of its spin axis, which is in turn determined by the YORP effect. Given long enough, these bodies will drift into regions where perturbations due to the gravity of the planets, especially those of Jupiter and Saturn, are enhanced. One such region, called a region of mean motion resonance, occurs where the orbital periods of two bodies are in a simple integer ratio. Resonances frequently act as dynamical escape hatches from the asteroid belt and are responsible for continually replenishing the population of near-Earth asteroids.

The distance traversed by fragments from an asteroid break-up can also, if carefully modelled, be used as a clock to determine when the collision took place[5]. Studying the Yarkovsky/YORP evolution could therefore also eventually allow us to explain the precise history of collision events in the asteroid belt over the past several billion years. Finally, the YORP effect can do extremely strange things in tandem with planetary perturbations. For example, consider the Koronis asteroid family: its known prograde-rotating bodies (that is, those spinning in the same direction in which they are moving) all have nearly identical spin rates, and have spin axes that roughly point towards the constellation Cassiopeia[11]. Rather than signifying some message from deep space, however, it seems that these asteroids were captured in special spin states after billions of years of similar YORP-driven evolution[12].

Thus, when it comes to asteroid dynamics, a light touch might be all that is needed to set big changes in motion. Although we may not be any closer to stopping the spinning up of those asteroids, we at least think we understand what's behind it. ■

William F. Bottke is in the Department of Space Studies, Southwest Research Institute, 1050 Walnut Street, Boulder, Colorado 80302, USA.
e-mail: bottke@boulder.swri.edu

1. Kaasalainen, M., Ďurech, J., Warner, B. D., Krugly, Y. N. & Gaftonyuk, N. M. *Nature* **446,** 420–422 (2007).
2. Lowry, S. C. *et al. Science* doi:10.1126/science.1139040 (2007).
3. Taylor, P. A. *et al. Science* doi:10.1126/science.1139038 (2007).
4. Rubincam, D. P. *Icarus* **148,** 2–11 (2000).
5. Bottke, W. F., Vokrouhlický, D., Rubincam, D. P. & Nesvorný, D. *Annu. Rev. Earth Planet. Sci.* **34,** 157–191 (2006).
6. Beekman, G. *J. Hist. Astron.* **37,** 71–86 (2006).
7. Chesley, S. R. *et al. Science* **302,** 1739–1742 (2003).
8. Paddack, S. J. Thesis, Catholic Univ. Am., Washington DC (1973).
9. Ostro, S. J. *et al. Science* **314,** 1276–1280 (2006).
10. Scheeres, D. J. *et al. Science* **314,** 1280–1283 (2006).
11. Slivan, S. M. *Nature* **419,** 49–51 (2002).
12. Vokrouhlický, D., Nesvorný, D. & Bottke, W. F. *Nature* **425,** 147–152 (2003).

ORGANIC CHEMISTRY

Synthesis undressed

John A. Porco Jr

Fragile chemical groups can be shielded from harsh reaction conditions by temporary protection. This approach is conventional wisdom for organic synthesis, but is it always the best solution?

Ask any organic chemist about their worst nightmare, and they'll probably give you the same answer: not being able to remove 'protecting groups'. These handy chemical groups are used to safeguard parts of a molecule that would otherwise be destroyed under the reaction conditions of a synthesis. The final step of a complex synthetic route often involves removing protecting groups to unmask the target compound. But if the wrong groups have been chosen, they might stay stubbornly attached to the molecule — or worse still, the molecule might be destroyed in the attempt to remove them. Years of work are frequently ruined in this way. On page 404 of this issue[1], Baran *et al.* show that complex, naturally occurring compounds can be constructed without using a single protecting group. This approach opens up avenues of research for discovering reactions that harness the intrinsic reactivity of unprotected organic molecules.

Protecting groups are ubiquitous in organic synthesis and often seem essential for performing multi-step sequences of reactions[2,3]. Not surprisingly, chemists have become reliant on these groups to access interesting molecules, especially naturally occurring compounds (generally referred to as natural products) with complex structures that are much loved as targets for synthesis[4]. But each group used adds two additional steps to any given reaction sequence — a protection step, where the groups are attached to the molecule, and a de-protection step, where they are removed. Because several different protecting groups are often used in one molecule, this can greatly decrease

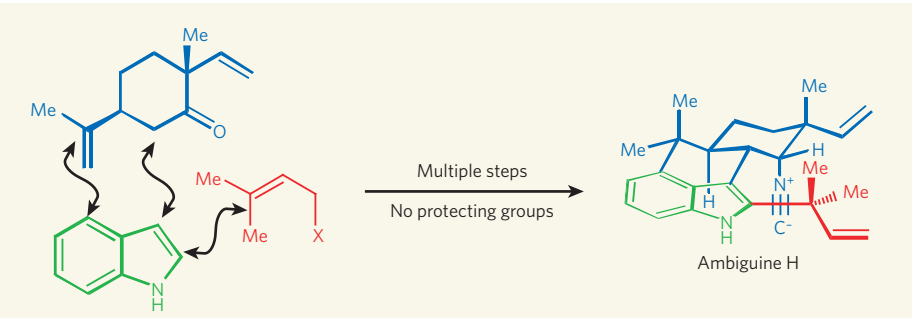


Figure 1 | Synthesis without protection. Baran *et al.*[1] have prepared ambiguine H, a naturally occurring compound originally extracted from marine organisms. Because they used no 'protecting groups' — temporary chemical groups that are attached to fragile parts of the molecule to shield them from destructive reaction conditions — the authors' synthesis is the shortest ever devised for this compound. The molecule is constructed from three fragments: an indole (green), a terpene (blue) and a prenyl unit (red). Curly arrows indicate where bonds are formed between the fragments; Me represents a methyl group (CH_3).

BEHAVIOURAL NEUROSCIENCE

Hare-brained flies

Using the fruitfly (pictured) as a model to investigate human traits such as attention span might seem odd. But the power of *Drosophila* genetics, together with previous studies pointing to sophisticated behavioural responses in this organism, in fact makes it an ideal choice for studying how our minds wander.

Bruno van Swinderen suspended flies in a cylindrical arena with rotating walls on which one of two simple visual stimuli was displayed (B. van Swinderen *Science* **315,** 1590–1593; 2007). He found that, each time the stimuli were switched, the fly's local field potential (LFP) activity — a measure of the total electrical activity at the junctions between neurons — increased. When the same object was displayed on both sides of the rotating cylinder there was no increased LFP response when it appeared anew. This ruled out the possibility that the elevated LFP response to a second object was simply due to a startle reflex.

When alternating the two visual stimuli, van Swinderen found that an interval of at least 50 seconds was required since the flies last saw the object for that stimulus to regain its novelty value, as measured by increased LFP activity. This response lasted an average of 9 seconds before the object lost its salience once more.

The author next performed these tests on two fly mutants — *dunce* and *rutabaga*. The proteins encoded by these genes normally alter levels of the same molecule, cAMP, and the mutants show similar defects in short-term memory. Van Swinderen found that the LFP activity in the brains of these mutants did not fluctuate appropriately in response to novel visual stimuli.

PHOTOTAKE/PHOTOLIBRARY.COM

Might it simply be that the general responsiveness to visual cues is defective in these mutants? Surprisingly, the answer seems to be no. Van Swinderen found not only that visual responsiveness was unaffected, but also that, following an initial delay, it was in fact far higher in the *dunce* mutants than in normal flies. Further experiments confirmed that although the mutant flies responded normally to visual stimuli, they were defective in identifying a new stimulus.

These results highlight the importance of the ability not just to pay attention, but also to divert it when necessary. The excessive responsiveness of the mutant flies to one visual stimulus seems to compromise short-term-memory activities such as shifting attention, or simultaneously paying attention, to another object. When the product of the *dunce* gene, the enzyme cAMP phosphodiesterase, was expressed throughout brain development, the attention defects associated with the mutation were overcome. This indicates that the developmental activity of cAMP is crucial for characteristics resembling attention in adult flies. Further analysis of both gene and enzyme should improve our understanding of what it takes to grab flies' — and perhaps humans' — attention.

Sadaf Shadan

the overall efficiency of a synthesis.

It wasn't always like this. Protection strategies simply weren't an option in the early days of organic chemistry, because many of the protecting groups widely used today were developed only in the past 50 years. Accordingly, some exemplary syntheses were devised that do not use protecting groups. For example, a useful method was developed in the early 1900s for converting sugars known as pentoses into larger sugars called hexoses[5]. Pentoses contain many reactive hydroxyl (OH) groups that chemists today would almost certainly protect. This three-step method has been invaluable for making sugars that are difficult to obtain from natural sources. More complex syntheses have also been reported. In 1957, a nine-step synthesis of muscarine — a natural product that mimics the action of certain neurotransmitters — did not use a single protecting group[6].

Another approach for avoiding protection strategies is to imitate biochemical routes found in nature. A landmark synthesis of this type was reported by Robert Robinson[7] in 1917; Robinson went on to win the Nobel Prize in Chemistry partly in recognition of this work. He prepared tropinone — a synthetic precursor of the drug atropine — in one step from simple starting materials, without protecting groups. This is considered to be an early example of a biomimetic cascade sequence[8], in which an initial reaction triggers a defined chain of other reactions, like dominoes toppling in a line. An extreme version of the biomimetic approach uses enzymes to mediate organic reactions[9,10]. Enzymes have evolved to work with naturally available molecules that clearly do not incorporate synthetic protecting groups.

But the early examples of syntheses free of protecting groups are not the norm. As time passed, the number of organic synthetic methods and chemical reagents expanded markedly. This led to an increased use of protecting groups in multi-step reaction sequences — so much so that this approach is hardly questioned today, and potentially interesting reactions that depend on the innate reactivity of unprotected molecules might be missed.

So Baran and co-workers[1] have gone back to basics. They describe synthetic routes to structurally complex compounds: ambiguine H and other related molecules that have been isolated from marine organisms. Each molecule is beautifully constructed by the authors without the use of any protecting groups. A good example of their approach is a step in which an aromatic, nitrogen-containing compound (an indole) reacts with another molecule known as a terpene (Fig. 1), mediated by a strong base and a copper salt[11]. This intriguing reaction relies on the presence of a hydrogen atom attached to the nitrogen of the indole fragment. But this hydrogen would not have been present if the nitrogen had been capped with a protecting group. After the introduction of an isonitrile group ($R–\overset{+}{N}\equiv\overline{C}$) into the terpene structure, the authors went on to append a hydrocarbon fragment (a prenyl group; Fig. 1) using an unorthodox reaction that relies again on the unprotected indole nitrogen, and on the innate reactivity of the isonitrile functional group. Remarkably, the syntheses of ambiguine H and related molecules required only 7–10 steps, in contrast to the 20–25 steps for previous routes to these syntheses that used protecting groups.

This study by Baran and co-workers shows that complex natural products can be synthesized more efficiently by reducing the use of protecting groups. But there are far greater implications of this work — the authors' strict avoidance of such groups will inspire new chemistry by exposing the intrinsic reactivity of chemical reagents and reactive intermediates. Further work on chemical synthesis using 'undressed' molecules should lead to many more advances and innovative developments in organic synthesis. And it might let chemists rest easy at night, safe in the knowledge that years of lab work won't be destroyed by recalcitrant protecting groups. ■

John A. Porco Jr is in the Department of Chemistry, Center for Chemical Methodology and Library Development, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA.
e-mail: porco@bu.edu

1. Baran, P. S., Maimone, T. J. & Richter, J. M. *Nature* **446,** 404–408 (2007).
2. Greene, T. W. & Wuts, P. G. M. *Protective Groups in Organic Synthesis* (Wiley-Interscience, New York, 2006).
3. Kocienski, P. J. *Protecting Groups* (Thieme, Stuttgart, 2003).
4. Hoffmann, R. W. *Synthesis* 3531–3541 (2006).
5. Lichtenthaler, F. W. *Angew. Chem. Int. Edn* **31,** 1541–1556 (1992).
6. Hardegger, E. & Lohse, F. *Helv. Chim. Acta* **40,** 2383–2389 (1957).
7. Robinson, R. *J. Chem. Soc. Trans.* **111,** 762–768 (1917).
8. Nicolaou, K. C., Montagnon, T. & Snyder, S. A. *Chem. Commun*. 551–564 (2003).
9. Whitesides, G. M. & Wong, C.-H. *Aldrichim. Acta* **16,** 27–34 (1983).
10. Koeller, K. M. & Wong, C.-H. *Nature* **409,** 232–240 (2001).
11. Baran, P. S. & Richter, J. M. *J. Am. Chem. Soc.* **126,** 7450–7451 (2004).

EVOLUTIONARY BIOLOGY

Adaptation under a microscope

Rosemary G. Gillespie and Brent C. Emerson

Experiments with microorganisms can guide thinking about the big questions being tackled by evolutionary biologists — for instance, how predation and immigration might play a role in adaptive radiation.

What accounts for biodiversity? Why do lineages of organisms diversify, and why are some lineages more species-rich than others? Charles Darwin was asking these questions 150 years ago, and we continue to do so today. Since Darwin, however, biologists have developed different theoretical and practical ways of tackling them, with experiments involving microbial systems featuring among the latter. Two examples of insights to emerge from such systems are provided by Meyer and Kassen[1] and Fukami *et al.*[2] elsewhere in this issue.

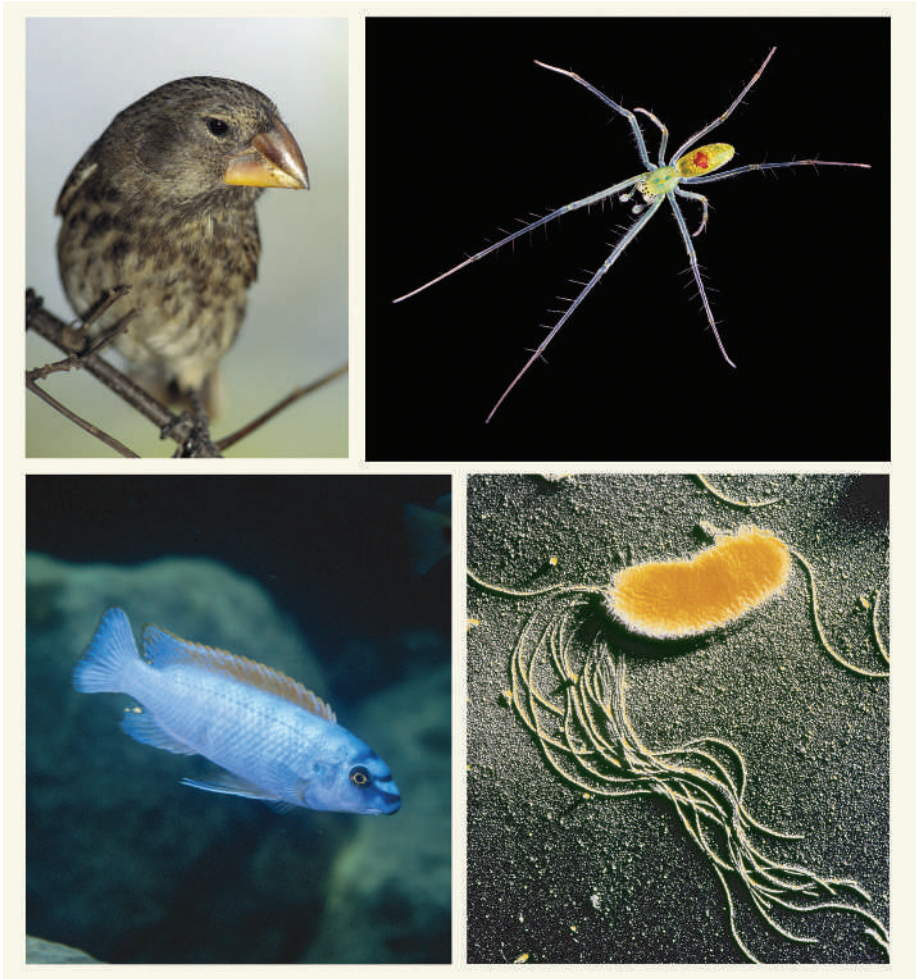
During the past century, the importance of repeated and successive iterations of adaptive radiation in the history of life has become increasingly apparent. Adaptive radiation is the rapid diversification of species to fill many ecological roles, and attempts to understand the factors involved have focused on isolated islands or lakes, where physical conditions would seem to favour this process. Well-known examples are the Darwin's finches of the Galapagos Islands[3] and the cichlid fishes of the large East African lakes[4] (Fig. 1). If immigration is constrained by geographical factors, species diversity can increase through evolution from a few colonists in response to open, often quite varied, ecological space. But evolution is generally a slow process, and biologists have had to rely on inference to elucidate the underlying mechanisms.

Hence the resort to microbial systems[5]. Evolution occurs more rapidly in such systems, allowing controlled experiments that provide insight into how communities develop through evolution or immigration, and the potential roles of competition and predation in driving the process. Particular use has been made of the soil bacterium *Pseudomonas fluorescens*. This bacterium exists as several different forms, or 'ecomorphs' — identifiable genotypes adapted to a particular niche — including SM (smooth), WS (wrinkly spreader) and FS (fuzzy spreader). Meyer and Kassen[1] and Fukami *et al.*[2] have used such microbial systems to investigate two factors that help to explain diversity through adaptive radiation: predation and immigration history.

Consider how diversity arises in communities of larger organisms: the assembly of species over ecological time involves immigration and extinction. However, given sufficient isolation, speciation — often in the form of adaptive radiation — may occur more rapidly than immigration, and be the primary contributor to species diversity. Speciation by adaptive radiation requires unoccupied ecological space, and sufficient geographical isolation to allow for genetic divergence. Moreover, the assembly of a community through adaptive radiation is expected to take much longer than that by immigration.

Accordingly, to understand the process of community assembly through evolution and adaptive radiation, inferences must be made from current ecological and morphological relationships among extant species within a lineage, and from the evolutionary history of that lineage. Historical approaches to investigate the causes of divergent selection that promote adaptive radiation have highlighted in particular the importance of competition between different species[6]. However, experiments showing a positive correlation between divergent selection and trait divergence in the presence of another factor, predation[7], support the old idea that predation can be a driver of adaptive evolutionary change. Yet, in the absence of other solid empirical evidence, the role of predation in promoting adaptive evolutionary change through diversifying selection has remained moot.

Meyer and Kassen[1] (page 432) use the *P. fluorescens* microbial system to provide this evidence. They examine the effect of a predator (a ciliated protozoan, *Tetrahymena thermophila*) on changes in diversity among bacterial genotypes. In the face of both competition and predation, bacterial ecomorphs show negative frequency-dependent selection — where the fitness of a genotype decreases with its frequency in the population — thus maintaining a diversity of ecomorphs. Frequency-dependent selection under competition seems to be caused by competition for resources, whereas under predation it is mediated by the refuge from predators provided by the floating bacterial mat. But perhaps the most exciting finding from these authors is that diversification of



M. MOFFETT/MINDEN PICTURES/FLPA

D. LITTSCHWAGER & S. MIDDLETON

A. BANNISTER/NHPA

T. BRAIN/SPL

Figure 1 | Bird, fish, spider, bacterium. Classic examples of adaptive radiation come from the Galapagos Islands (Darwin's finches) and the lakes of East Africa (cichlid fishes). Hawaii provides further instances with the tetragnathid spiders. Evolution is generally slow in natural ecosystems, however, hence the use of experimental systems featuring bacteria such as *Pseudomonas fluorescens*[1,2].

bacteria is delayed in the presence of predators. The reason seems to be that predation reduces the intensity of resource competition — and hence diversifying selection — among bacterial ecomorphs. The results suggest that predation may play a prominent, but often unnoticed, role in adaptive radiation.

Another challenge to inference-based analyses of adaptive radiation has been immigration, and the intractability of determining the importance of the sequence in which different species arrive in an isolated habitat. Arrival order may have an effect not only on whether a lineage diversifies in the first place, but also on the eventual species composition in a given community that develops over evolutionary time.

Fukami *et al.*[2] (page 436) have used the *P. fluorescens* system to demonstrate the importance of immigration history in dictating the eventual composition of diversity in a community. Their study shows that the SM ecomorph of *P. fluorescens*, if left on its own, evolves predictably as noted above to form one FS ecomorph and multiple WS ecomorphs. But they observed that small differences in the timing of immigration markedly affected the eventual diversity in a community: if the WS ecomorph was also introduced, they found that by controlling when this specialist ecomorph arrived, it could suppress diversification altogether. These results support data from studies of macroecological communities that document differences in the sets of ecomorphs arising in different situations and the dynamic nature of community assembly over evolutionary time[8,9].

Both of these studies[1,2] contribute to our understanding of the historical contingencies of community assembly. Meyer and Kassen's work[1] highlights a role for interactions among taxa in promoting evolutionary diversification. This supports the view that taxa in species-rich communities may undergo more evolutionary change than do those in less species-rich communities[10]. Analyses of island species show parallels between the formation of communities through evolutionary processes and those formed over ecological time — highlighting the notion that evolution is nothing but ecology writ large[11]. Interestingly, some communities lack the full suite of potential niche specialists[9], and the results of Fukami *et al.*[2] raise the possibility that inconsistencies are partly due to immigration history. 'First come, first served' seems to hold when it comes to filling empty ecological space. The challenge is to apply the knowledge gained from these rich bacterial systems to a more general appreciation of adaptive radiation and global patterns of biodiversity. ■

Rosemary G. Gillespie is in the Department of Environmental Science, University of California, Berkeley, California 94720, USA.
Brent C. Emerson is in the Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.
e-mails: gillespie@berkeley.edu; b.emerson@uea.ac.uk

1. Meyer, J. R. & Kassen, R. *Nature* **446,** 432–435 (2007).
2. Fukami, T., Beaumont, H. J. E., Zhang, X.-X. & Rainey, P. B. *Nature* **446,** 436–439 (2007).
3. Grant, P. R. *Ecology and Evolution of Darwin's Finches* (Princeton Univ. Press, 1999).
4. Stiassny, M. L. J. & Meyer, A. *Sci. Am.* **280,** 64–69 (1999).
5. Rainey, P. B. & Travisano, M. *Nature* **394,** 69–72 (1998).
6. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
7. Nosil, P. & Crespi, B. J. *Proc. Natl Acad. Sci. USA* **103,** 9090–9095 (2006).
8. Losos, J. B. *Phil. Trans. R. Soc. Lond. B* **349,** 69–75 (1995).
9. Gillespie, R. G. *Science* **303,** 356–359 (2004).
10. Emerson, B. C. & Kolm, N. *Nature* **434,** 1015–1017 (2005).
11. Van Valen, L. *Science* **180,** 488 (1973).

BIOCHEMISTRY

Molecular cannibalism

Steven E. Ealick and Tadhg P. Begley

The biosynthesis of vitamin B_{12} has fascinated generations of scientists, but part of the pathway was unknown. The missing enzymatic link has now been found, only to raise more mechanistic questions.

Most people who take vitamin supplements are unaware of the scientific history behind their unassuming tablets. Vitamin B_{12} is an excellent case in point — no less than four Nobel prizes have been awarded for work relating to this seemingly commonplace compound (Box 1, overleaf). The biosynthesis of vitamin B_{12} is an integral part of this scientific heritage, so one could be forgiven for thinking that there is nothing left to discover. But this is not so. The origins of one fragment of this vitamin, known as the dimethylbenzimidazole (DMB) ligand, have remained an enigma. On page 449 of this issue, Taga *et al.*[1] finally unravel the mystery by identifying the enzyme responsible for DMB biosynthesis, and describing its structure.

Vitamin B_{12} is essential for human health — its absence leads to the autoimmune disease known as pernicious anaemia. It is perhaps surprising to learn that only bacteria, fungi and algae produce this vitamin, whereas animals and plants must obtain it from their diet. The term 'vitamin B_{12}' is actually a general description for several structurally related compounds, two of which are the major biologically active variants. The first of these is methyl cobalamin,

50 YEARS AGO

It is curious how few facts of real importance are known about the life and parentage of Archimedes, while the trivial story of his leaping out of his bath shouting "Heureka" is familiar to every schoolboy. The first record of it, however, is in the works of Vitruvius, written about two hundred years after Archimedes's death, so that there was ample time for the story to have been embroidered, even if it is not a pure invention. It is much the same with the account of his launching a large ship single-handed, saying, "Give me a place to stand on and I will move the earth", and with the myth that he burned the Roman fleet by using mirrors on a sunny day. These traditional stories, and others, are critically considered in Prof. E. J. Dijksterhuis's book.

From *Nature* 23 March 1957.

100 YEARS AGO

Nature Knowledge in Modern Poetry — In this book the author deals in a very interesting manner with the many references to the aspects of nature in the poetical works of Tennyson, Wordsworth, Matthew Arnold, and Lowell... Interest in the insect world is shown to a greater extent by Tennyson, for he alludes to it frequently, and always with the accuracy which reveals great knowledge... Tennyson's love of geology is apparent in the frequent references to it and the similes he gives, which clearly show he must have read a good deal on this as indeed on many other less popular subjects; for instance, he does not shun allusions to the nebular hypothesis, spectrum analysis, and astronomy. It seems evident that he accepted the theory of evolution, for many quotations might be made to show it...

"Evolution ever climbing after some ideal good,
And reversion ever dragging Evolution in the mud."

From *Nature* 21 March 1907.

50 & 100 YEARS AGO

which functions as a methyl (CH_3) donor in the biosynthesis of the amino acid methionine. The second is 5-deoxyadenosylcobalamin ($AdoB_{12}$), which is the source of a free radical — the adenosyl radical — that catalyses various reactions that are triggered by the removal of a hydrogen atom from a compound.

Three generations of biological chemists have been captivated by $AdoB_{12}$ because of its complexity and novelty; until recently, it was the only known example of an organometallic compound used by living systems. The biosynthesis of $AdoB_{12}$ took the best part of 40 years to unravel because of its complexity, but the process is now relatively well understood[2]. Only the construction of DMB was unresolved, although it was known to be formed by the destruction of vitamin B_2 (riboflavin)[3]. Such molecular cannibalization of one vitamin to form another is an emerging theme in vitamin biosynthesis and has also been observed in the biosyntheses of vitamin H, vitamin B_1 and lipoic acid[4–6]. Experiments with isotopically labelled riboflavin have shown where the atoms from vitamin B_2 end up in DMB, but the enzyme (referred to generically as a flavin destructase) that mediates this process was unknown[3].

Taga *et al.*[1] report that an enzyme called BluB, which they studied in the bacterium *Sinorhizobium meliloti*, is the missing flavin destructase, and they go on to describe its structure. The authors also characterize an early intermediate in the reaction pathway in which molecular oxygen is poised to attack the flavin (flavin mononucleotide; FMN). The information obtained from this structure provides detailed data on how the enzyme uses oxygen to achieve the breakdown of riboflavin to give DMB.

The mechanism of this remarkable reaction has not yet been established, although Taga *et al.*[1] propose two possibilities. The authors acknowledge that these suggestions have their limitations, however. The first mechanism requires a deprotonation by an amino-acid residue that is unlikely to be a strong enough base to do the job. The second mechanism involves a reduction by a negatively charged hydrogen ion (a hydride, H^-) that is derived from a hydroxyl group (OH) on the substrate; this is without precedent, because the hydrogen atoms in hydroxyl groups form positive ions rather than hydrides.

The complexity of the chemistry required for the conversion of FMN to DMB is in sharp contrast to the relative simplicity of the BluB active site. This simplicity suggests to us an alternative mechanism in which a series of reactions with water dismantles the flavin after it has reacted with molecular oxygen (Fig. 1). Close relatives of all of the reactions in this proposal have been found in other systems[7], and would be more consistent with the structure of the BluB active site.

It seems that in answering one question Taga and colleagues have raised many more. The availability of a robust biological system for producing BluB, along with the enzyme's structure, sets the stage for the complete mechanistic characterization of this unusual reaction. ■

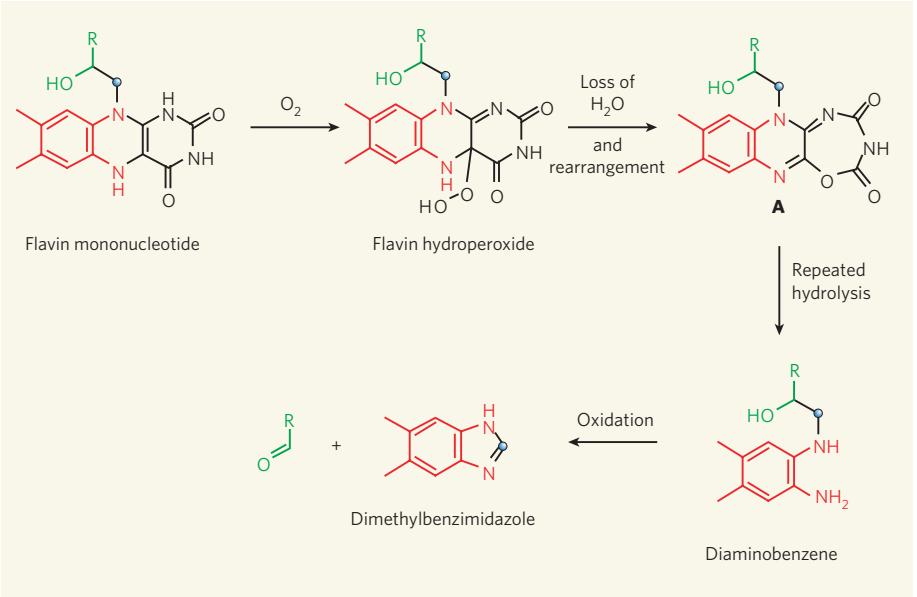


Figure 1 | A possible mechanism for the biosynthesis of a vitamin B_{12} fragment. Taga *et al.*[1] show that the BluB enzyme catalyses a reaction in which reduced flavin mononucleotide (FMN) is converted into dimethylbenzimidazole (DMB), a fragment of vitamin B_{12}. In the mechanism proposed here, FMN reacts with oxygen to yield flavin hydroperoxide, which then undergoes a ring expansion with loss of a water molecule to produce intermediate **A**. This disintegrates further as a result of four hydrolysis reactions in which it is attacked by water molecules. The resulting diaminobenzene product undergoes two oxidation reactions, to form DMB[7]. R represents the side chain of FMN; the blue circle marks the carbon atom in the side chain that is incorporated in DMB.

Box 1 | The illustrious history of vitamin B_{12}

Work related to vitamin B_{12} has resulted in four Nobel prizes. In the 1920s, George Minot showed that including large amounts of raw liver in the diets of anaemic dogs cured their condition. Inspired by this work, William Murphy and George Whipple isolated the curative substance and named it vitamin B_{12}. The three were awarded the 1934 Nobel Prize in Physiology or Medicine for their discoveries.

In the 1950s and early 1960s, Dorothy Hodgkin worked out the molecular structure of vitamin B_{12} using X-ray crystallography[8]. The structural analysis of such a complex molecule had never been attempted before, and she was awarded the 1964 Nobel Prize in Chemistry for this accomplishment. With the structure of the vitamin in hand, the chemists Robert Woodward and Albert Eschenmoser set out to synthesize it. The total synthesis took more than ten years to complete and led in part to the 1965 Nobel Prize in Chemistry being awarded to Woodward[9].

A crucial step in the complex synthesis of vitamin B_{12} provided the impetus for Woodward and the theoretical chemist Roald Hoffmann to develop a method for predicting which kinds of chemical reaction are favoured. The method applies to a class of chemical reactions known as pericyclic reactions. The resulting set of rules, which state that the favoured products are those in which orbital symmetry is conserved, are now known as the Woodward–Hoffmann rules. A similar set of rules was independently developed by the physical chemist Kenichi Fukui. Both Fukui and Hoffmann were awarded the 1981 Nobel Prize in Chemistry for their insight; had Woodward not died two years earlier, he would almost certainly have shared this prize. **S.E.E. & T.P.B.**

Steven E. Ealick and Tadhg P. Begley are in the Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, USA.
e-mails: see3@cornell.edu; tpb2@cornell.edu

1. Taga, M. E., Larsen, N. A., Howard-Jones, A. R., Walsh, C. T. & Walker, G. C. *Nature* **446,** 449–453 (2007).
2. Roessner, C. A., Santander, P. J. & Scott A. I. *Vitam. Horm.* **61,** 267–297 (2001).
3. Renz, P. in *Chemistry and Biochemistry of B_{12}* (ed. Banerjee, R.) 557–575 (Wiley, Weinheim, 1999).
4. Berkovitch, F., Nicolet, Y., Wan, J. T., Jarrett, J. T. & Drennan, C. L. *Science* **303,** 76–80 (2004).
5. Cicchillo, R. M. & Booker, S. J. *J. Am. Chem. Soc.* **127,** 2860–2861 (2005).
6. Chatterjee, A., Jurgenson, C. T., Schroeder, F. C., Ealick, S. E. & Begley, T. P. *J. Am. Chem. Soc.* **128,** 7158–7159 (2006).
7. Maggio-Hall, L. A., Dorrestein, P. C., Escalante-Semerena, J. C. & Begley, T. P. *Org. Lett.* **5,** 2211–2213 (2003).
8. Hodgkin, D. C. *et al. Proc. R. Soc. Lond. A* **242,** 228–263 (1957).
9. Eschenmoser, A. *Angew. Chem.* **100,** 5–40 (1988).

bacteria is delayed in the presence of predators. The reason seems to be that predation reduces the intensity of resource competition — and hence diversifying selection — among bacterial ecomorphs. The results suggest that predation may play a prominent, but often unnoticed, role in adaptive radiation.

Another challenge to inference-based analyses of adaptive radiation has been immigration, and the intractability of determining the importance of the sequence in which different species arrive in an isolated habitat. Arrival order may have an effect not only on whether a lineage diversifies in the first place, but also on the eventual species composition in a given community that develops over evolutionary time.

Fukami *et al.*[2] (page 436) have used the *P. fluorescens* system to demonstrate the importance of immigration history in dictating the eventual composition of diversity in a community. Their study shows that the SM ecomorph of *P. fluorescens*, if left on its own, evolves predictably as noted above to form one FS ecomorph and multiple WS ecomorphs. But they observed that small differences in the timing of immigration markedly affected the eventual diversity in a community: if the WS ecomorph was also introduced, they found that by controlling when this specialist ecomorph arrived, it could suppress diversification altogether. These results support data from studies of macroecological communities that document differences in the sets of ecomorphs arising in different situations and the dynamic nature of community assembly over evolutionary time[8,9].

Both of these studies[1,2] contribute to our understanding of the historical contingencies of community assembly. Meyer and Kassen's work[1] highlights a role for interactions among taxa in promoting evolutionary diversification. This supports the view that taxa in species-rich communities may undergo more evolutionary change than do those in less species-rich communities[10]. Analyses of island species show parallels between the formation of communities through evolutionary processes and those formed over ecological time — highlighting the notion that evolution is nothing but ecology writ large[11]. Interestingly, some communities lack the full suite of potential niche specialists[9], and the results of Fukami *et al.*[2] raise the possibility that inconsistencies are partly due to immigration history. 'First come, first served' seems to hold when it comes to filling empty ecological space. The challenge is to apply the knowledge gained from these rich bacterial systems to a more general appreciation of adaptive radiation and global patterns of biodiversity. ■

Rosemary G. Gillespie is in the Department of Environmental Science, University of California, Berkeley, California 94720, USA.
Brent C. Emerson is in the Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.
e-mails: gillespie@berkeley.edu; b.emerson@uea.ac.uk

1. Meyer, J. R. & Kassen, R. *Nature* **446,** 432–435 (2007).
2. Fukami, T., Beaumont, H. J. E., Zhang, X.-X. & Rainey, P. B. *Nature* **446,** 436–439 (2007).
3. Grant, P. R. *Ecology and Evolution of Darwin's Finches* (Princeton Univ. Press, 1999).
4. Stiassny, M. L. J. & Meyer, A. *Sci. Am.* **280,** 64–69 (1999).
5. Rainey, P. B. & Travisano, M. *Nature* **394,** 69–72 (1998).
6. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
7. Nosil, P. & Crespi, B. J. *Proc. Natl Acad. Sci. USA* **103,** 9090–9095 (2006).
8. Losos, J. B. *Phil. Trans. R. Soc. Lond. B* **349,** 69–75 (1995).
9. Gillespie, R. G. *Science* **303,** 356–359 (2004).
10. Emerson, B. C. & Kolm, N. *Nature* **434,** 1015–1017 (2005).
11. Van Valen, L. *Science* **180,** 488 (1973).

BIOCHEMISTRY

Molecular cannibalism

Steven E. Ealick and Tadhg P. Begley

The biosynthesis of vitamin B_{12} has fascinated generations of scientists, but part of the pathway was unknown. The missing enzymatic link has now been found, only to raise more mechanistic questions.

Most people who take vitamin supplements are unaware of the scientific history behind their unassuming tablets. Vitamin B_{12} is an excellent case in point — no less than four Nobel prizes have been awarded for work relating to this seemingly commonplace compound (Box 1, overleaf). The biosynthesis of vitamin B_{12} is an integral part of this scientific heritage, so one could be forgiven for thinking that there is nothing left to discover. But this is not so. The origins of one fragment of this vitamin, known as the dimethylbenzimidazole (DMB) ligand, have remained an enigma. On page 449 of this issue, Taga *et al.*[1] finally unravel the mystery by identifying the enzyme responsible for DMB biosynthesis, and describing its structure.

Vitamin B_{12} is essential for human health — its absence leads to the autoimmune disease known as pernicious anaemia. It is perhaps surprising to learn that only bacteria, fungi and algae produce this vitamin, whereas animals and plants must obtain it from their diet. The term 'vitamin B_{12}' is actually a general description for several structurally related compounds, two of which are the major biologically active variants. The first of these is methyl cobalamin,

50 YEARS AGO

It is curious how few facts of real importance are known about the life and parentage of Archimedes, while the trivial story of his leaping out of his bath shouting "Heureka" is familiar to every schoolboy. The first record of it, however, is in the works of Vitruvius, written about two hundred years after Archimedes's death, so that there was ample time for the story to have been embroidered, even if it is not a pure invention. It is much the same with the account of his launching a large ship single-handed, saying, "Give me a place to stand on and I will move the earth", and with the myth that he burned the Roman fleet by using mirrors on a sunny day. These traditional stories, and others, are critically considered in Prof. E. J. Dijksterhuis's book.

From *Nature* 23 March 1957.

100 YEARS AGO

Nature Knowledge in Modern Poetry — In this book the author deals in a very interesting manner with the many references to the aspects of nature in the poetical works of Tennyson, Wordsworth, Matthew Arnold, and Lowell... Interest in the insect world is shown to a greater extent by Tennyson, for he alludes to it frequently, and always with the accuracy which reveals great knowledge... Tennyson's love of geology is apparent in the frequent references to it and the similes he gives, which clearly show he must have read a good deal on this as indeed on many other less popular subjects; for instance, he does not shun allusions to the nebular hypothesis, spectrum analysis, and astronomy. It seems evident that he accepted the theory of evolution, for many quotations might be made to show it...

"Evolution ever climbing after some ideal good,
And reversion ever dragging Evolution in the mud."

From *Nature* 21 March 1907.

50 & 100 YEARS AGO

OBITUARY

Alan Graham MacDiarmid (1927–2007)

Pioneer of conducting polymers, and proud Antipodean.

Alan MacDiarmid, who died on 7 February, was a pioneer of the field of intrinsically conducting polymers. He was born on 14 April 1927 in Masterton on the North Island of New Zealand. His Nobel autobiography describes a relatively frugal upbringing in a loving and generous family. When extra guests were invited back to the MacDiarmid home for a welcoming meal, the family was instructed to "FHB" — family hold back — while the guests were fed first. This isolated and caring Antipodean environment shaped a generosity of spirit in MacDiarmid that was appreciated by all who came into close contact with him.

MacDiarmid left high school aged 16 and put himself through the then Victoria University College, Wellington, studying part-time while working as a lab boy. A Fulbright scholarship to the University of Wisconsin followed, where he completed a PhD under Norris Hall, and then won a New Zealand Shell scholarship to work on silicon hydrides with Harry Emeléus at the University of Cambridge, UK. With characteristic modesty, MacDiarmid described this opportunity as one he could not miss, even though it meant his studying for a second PhD. The terms of the scholarship also demanded that he remained single; eventually, however, he was able to marry Marian Mathieu in the chapel of Sidney Sussex College, Cambridge.

After a brief spell at the University of St Andrews, UK, MacDiarmid entered the chemistry department at the University of Pennsylvania in Philadelphia in 1964, where he remained essentially until his death. There, he met the condensed-matter physicist Alan Heeger, who suggested a collaboration on conductors that he described as SNX. These were inorganic covalent materials containing sulphur and nitrogen, $(SN)_x$. MacDiarmid was initially unimpressed, assuming that everyone would know that metallic tin, $(Sn)_x$, was conducting.

The encounter led to collaboration and lifelong friendship, culminating in the award of the 2000 Nobel Prize in Chemistry to Heeger, MacDiarmid and Hideki Shirakawa. Shirakawa's involvement was serendipitous: MacDiarmid had presented a seminar on conducting $(SN)_x$ materials at the Tokyo Institute of Technology in 1975, which Shirakawa did not attend because $(SN)_x$ was not his area, and he had missed the 'conducting' qualifier. Meeting after the seminar, Shirakawa showed MacDiarmid a silvery film that he had prepared by polymerizing acetylene, C_2H_2, in a toluene solvent with differing ratios of catalyst. Realizing the significance of this 'polyacetylene', MacDiarmid invited Shirakawa back to Philadelphia, soliciting funding support from Kenneth Wynne at the US Office of Naval Research. This funding decision was surely one of the most far-sighted of the period.

Together at Pennsylvania, MacDiarmid and Shirakawa first investigated improving the conductivity of polyacetylene by washing out impurities — but the purer the polymer became, the lower was its conductivity. They next tried to modify the polyacetylene with bromine, as had been done with $(SN)_x$, and observed an increase in conductivity. But it was when, in collaboration with Heeger and his group, they tried iodine that the real surprise came. Astoundingly, the iodine-doped polyacetylene showed a conductivity of 30 siemens per centimetre, an increase of seven orders of magnitude over the undoped material. The first conducting polymer had been produced.

Until that point, polymers had been regarded as excellent insulators, and the work changed the way scientists thought about plastics. Enormous interest in conducting and semiconducting polymers immediately followed. Improved and elegant precursor routes to polyacetylene were developed, and its conductivity was increased to a level almost as good as that of metallic copper, around 10^5 S cm^{-1}. All sorts of optoelectronic applications — transistors, solar cells and polymer light-emitting devices — have been the result.

When Heeger moved to Santa Barbara in 1982, MacDiarmid continued fruitful collaborations with many others, opening up the field of polyanilines and related stable, soluble conducting polymers with controllable properties. During his scientific career, he influenced many people with his enthusiastic and enquiring approach, and his generosity of spirit. That warmth extended not just to his close collaborators: he was always to be seen at conferences asking a speaker or a student presenting a poster an encouraging question that made them feel good about their work.

Although he lived and worked in the United States, Alan MacDiarmid treasured his Antipodean roots. He loyally supported science throughout the Asia–Pacific region: in New Zealand, where his achievements were honoured with the naming of the MacDiarmid Institute for Advanced Materials and Nanotechnology in Wellington; in Australia, where he chaired the international advisory committee for the University of Wollongong's Centre of Excellence in Electromaterials Science; and in China, through the MacDiarmid Laboratory at Jilin University. In the last four years of his life, he held two positions, at Pennsylvania and the University of Texas, Dallas, where his attention increasingly turned to climate change, and particularly the use of biofuels.

L. KANE-MAGUIRE

Alan enjoyed life to the full. At the biennial International Conference on Synthetic Metals, of which he was a founding member, and at many other international events, he could be relied on to demonstrate his version of the *haka*, an indigenous New Zealand war dance. MacDiarmid — pictured here with his Nobel medal at the University of Wollongong in 2001 — was a recipient of many honours in his later life. He was appointed as one of the only 20 concurrently living members of the Order of New Zealand in 2001, and was a member of the US National Academy of Sciences and a fellow of the Royal Society of London. In 2002, he received an honorary doctorate from the University of Cambridge, and was inducted an honorary fellow of Sidney Sussex College in the chapel where he was first married.

Marian, Alan's wife from that marriage, died in 1990. After a 14-year courtship he was married again, to Gayl Gentile, with whom many of Alan's scientific colleagues have also become affectionate friends. He is survived by Gayl, three daughters and a son from his first marriage, and nine grandchildren.

Andrew Holmes

Andrew Holmes is at the Bio21 Institute, University of Melbourne, Building 102, 30 Flemington Road, Parkville, Victoria 3010, Australia.
e-mail: aholmes@unimelb.edu.au

Q&A

ORGANOMETALLIC CHEMISTRY

C–H activation

Robert G. Bergman

The stability of the chemical bonds in saturated hydrocarbons makes them generally unreactive. But the invention of processes in which carbon–hydrogen (C–H) bonds in hydrocarbons can be activated is allowing chemists to exploit organic compounds in previously unimaginable ways.

Why is C–H activation such big news?

These reactions could revolutionize the chemical industry. Natural gas, for example, is best known as a fuel. But it is also a vast, low-cost feedstock of hydrocarbons that remains untapped as a raw material, simply because there has been no easy way of turning it into synthetically useful compounds. This could be about to change. Methane — CH_4, the main constituent of natural gas — can now be converted directly into derivatives of methanol (CH_3OH) in a high-yielding C–H activation reaction that is catalysed by metal salts in solution. Methanol is the starting point for many industrial processes, and is ultimately incorporated into such products as plastics and paints. C–H activation also allows chemical groups to be placed directly in a molecule where none existed before, a process that previously often needed several steps. This is especially useful for shortening multi-step syntheses, which are commonly used in drug discovery.

What stimulated chemists to work on C–H activation?

C–H bonds are ubiquitous in organic molecules, yet many of these could not be exploited for chemical reactions. A few methods were available, but these were unselective — they yielded a complicated mixture of products. So the uncharted territory of C–H bond reactivity was an irresistible area for chemists to explore. More pragmatically, it was, perhaps, the realization that hydrocarbon feedstocks are inefficiently used. Organic molecules, based on carbon chains, are the basis of most of the non-metallic materials found in everyday life. Plastics, for example, have replaced metals and ceramics in many applications. These organic materials are made by chemical synthesis from the hydrocarbons found in petroleum. But there's a problem: only about half of the hydrocarbons are reactive enough to take part in traditional chemical reactions. C–H activation gets around this by stimulating inert hydrocarbons to react with other molecules.

Why are only half of the hydrocarbon feedstocks chemically reactive?

The answer lies in the stability of their chemical bonds. Hydrocarbons are simple molecules that contain only carbon–carbon (C–C) and carbon–hydrogen (C–H) bonds (Fig. 1). There are two general classes: saturated hydrocarbons (alkanes), which contain only single bonds, and unsaturated hydrocarbons (arenes, alkenes and alkynes), which also contain some C–C multiple bonds. The single bonds in hydrocarbons are very stable, which is why alkanes are quite inert, although they can be slowly oxidized, unselectively, in air. But the C–C multiple bonds in unsaturated hydrocarbons are far less stable and so are more prone to chemical attack than are C–C or C–H single bonds.

How are hydrocarbons converted into more useful compounds?

Many reagents have been developed that introduce synthetically useful chemical groups into unsaturated hydrocarbons. At one time, these reagents were primarily other organic compounds, or were based on elements from the main groups of the periodic table. But some of the best modern reagents are based on transition metals such as iron, osmium, palladium, rhodium and ruthenium. This has created a research boom in organometallic chemistry — the interdisciplinary area that lies between organic and inorganic chemistry.

So what's the problem?

There are two. First, if a chemist wants to perform a reaction with an alkane, he or she is forced to 'activate' single bonds (either C–C or C–H) in the molecule, so that they become prone to chemical attack. Second, there is very little difference in reactivity between the various C–H bonds in alkanes, so targeting a specific C–H bond is difficult. Currently, very few reagents can activate specific C–H bonds in alkanes under mild reaction conditions.

What exactly is meant by 'activation'?

Most generally, this means treating a C–H bond in some way that allows a reagent to react

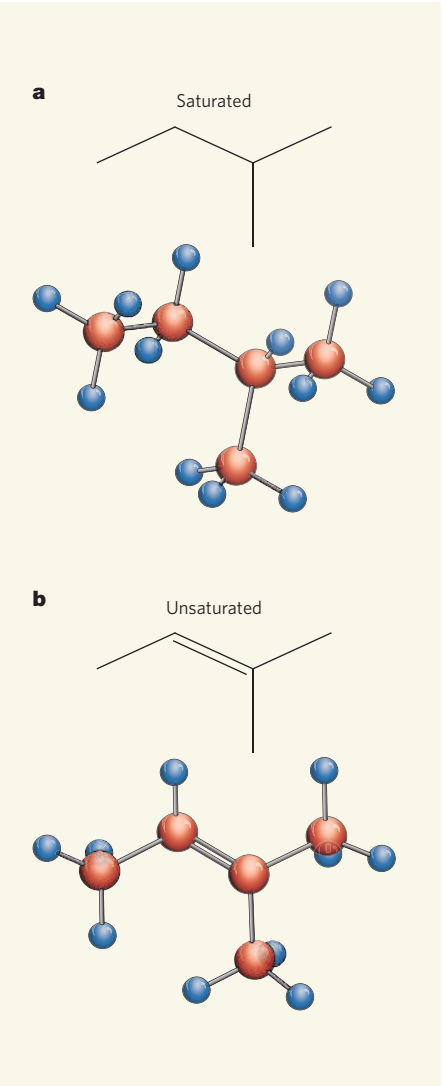


Figure 1 | Saturation in hydrocarbons. **a**, Saturated hydrocarbons are usually represented as shown in the top structure. The ball-and-stick model underneath shows that they consist of carbon (red) and hydrogen atoms (blue) connected by single bonds. **b**, Unsaturated hydrocarbons also contain some double bonds, which are more reactive than single bonds and make unsaturated hydrocarbons more reactive than saturated ones.

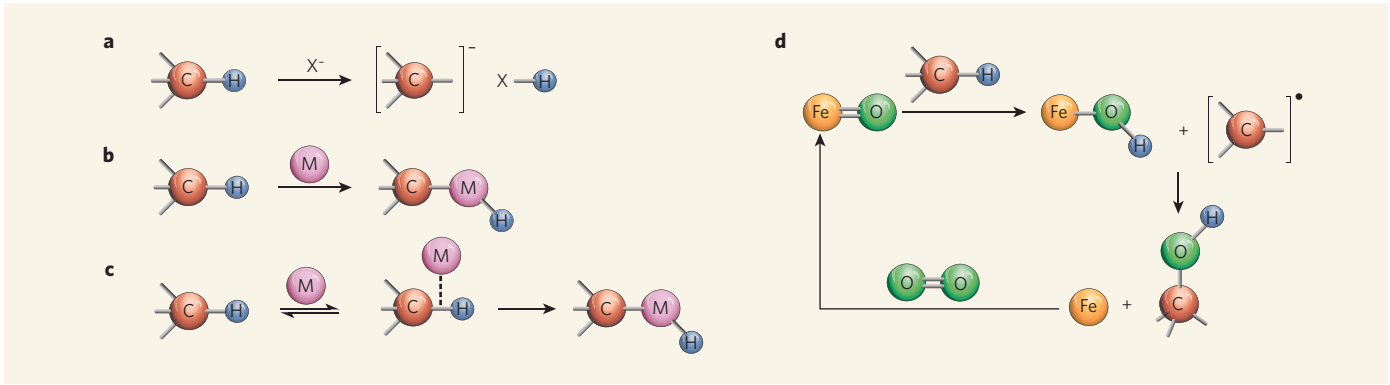


Figure 2 | Mechanisms for C–H activation reactions. **a**, Classical methods for C–H activation reactions simply removed the hydrogen atom, using either a base (X^-, as shown here), a free radical or an electrophile (reactions for the last two reagent types are not shown for simplicity). **b**, In oxidative addition reactions, a metal atom (M) inserts itself between the atoms of the C–H bond. **c**, The latest data suggest that oxidative additions proceed in two steps. A metal binds weakly to the C–H bond in an equilibrium process, before inserting itself between the atoms of the bond. **d**, The active sites of certain enzymes contain iron complexes in which iron (Fe) is bound to oxygen with a double bond. These might react with C–H bonds to yield a carbon radical (indicated by the dot) and an intermediate iron complex containing a hydroxyl group (OH). If the hydroxyl group transfers to the carbon atoms, the remaining iron complex could then be oxidized back to its original form, ready to repeat the cycle with another C–H bond. Atoms are not drawn to scale.

rapidly with the carbon atom. A second reaction then yields a stable product in which the C–H bond is replaced by a new bond, C–X, where X is often a nitrogen, oxygen or carbon atom in another molecule. The classical methods for this simply involved removing the hydrogen, possibly along with one or two of the bonding electrons (Fig. 2a). But these methods usually required a chemical group to be adjacent to the targeted C–H bond — saturated hydrocarbons, by definition, possess no such chemical groups. Alternatively, 'brute-force' reagents were required, such as strong acids or bases, or reactive free radicals; this compromised the selectivity of the reactions and prevented them from being performed under mild conditions.

What was the solution to these problems?

The solution was a new kind of reaction, reported in the late 1960s (Box 1). In these so-called oxidative addition reactions, a metal from a soluble metal salt or complex seems to insert itself between the atoms of a C–H bond, yielding an unstable product that contains a carbon–metal bond and a metal–hydrogen bond (Fig. 2b). The oxidation state of the metal increases during the transformation, hence the name oxidative addition. The latest evidence suggests that oxidative addition is not a simple one-step process. Instead, it seems to proceed through an intermediate in which the C–H bond is weakly bound to the metal centre, but where the C–H bond is not yet broken (Fig. 2c). The idea that a simple C–H bond, which does not have electrons that are readily available for binding to metals, could behave in this way is unprecedented. An understanding of such mechanistic details is crucial for developing the next generation of C–H activation reactions.

Are there any other methods for activating C–H bonds?

Another possible route uses solid metallic materials — known as heterogeneous catalysts — that can also react with C–H bonds. These reactions are more complicated than simple oxidative addition reactions (which take place in solution) and the mechanisms are more difficult to identify by spectroscopic techniques because the catalysts don't dissolve. This method of C–H activation has been studied extensively by physical chemists and chemical engineers, and constitutes a separate subject in its own right.

When did C–H activation really take off?

The field took off during the 1980s, when there was a dramatic increase in the number of metal salts and complexes that were found to initiate C–H activation by oxidative addition. But the drawback was that most of these transformations required equal amounts, in moles, of the hydrocarbon and the metal, and both partners were consumed during the reaction. This is not acceptable for large-scale chemistry, as the metals involved are generally more expensive than the products.

What is being done to improve the chemistry?

There has been an explosion of interest in the use of catalytic reactions for bringing about oxidative addition for C–H activation. In these catalytic processes, the oxidative addition product is a transient intermediate that immediately reacts with other reagents to introduce a new atom. The catalytic metal is released so that it can attack another molecule of hydrocarbon. Because the metal is not consumed, such reactions often need only tiny amounts of catalyst. An excellent example is the recent discovery that rhodium catalysts directly convert the C–H bonds at the ends of alkane chains into carbon–boron bonds; the products of such reactions are very useful for synthetic organic chemistry. Another example is the discovery that methane can be converted into methanol derivatives with unusually high yields using platinum complexes in strong acid solution. Catalytic reactions are especially important in nature, where enzymes (many of which bear metals at their active sites) mediate a wide range of biological transformations.

Are there any enzymatic C–H activation reactions?

Yes, some very important ones. Cytochrome P450 enzymes typically catalyse the conversion of C–H bonds to C–O bonds in organic compounds. In humans, these enzymes are involved in making cholesterol, steroids and other lipids; they also metabolize drugs, converting them to highly oxidized compounds that can be excreted by the body. The active sites of these enzymes contain iron atoms that play a crucial role in the C–H activation process. Another striking example of a C–H activating enzyme is methane monooxygenase, which was recently discovered in a class of bacterium that lives at the interface of aerobic and anaerobic environments. This enzyme converts methane to methanol, although it can also oxidize several other organic compounds.

Is there anything we can learn from these enzymes?

They have already shown us a completely different approach to C–H activation. The enzymes mentioned above have iron (Fe) ions at their active sites. Current evidence suggests that these iron centres react with oxygen to make a highly reactive iron–oxygen double bond, Fe=O (Fig. 2d). The critical activation step involves the addition of the atoms in a C–H bond to the Fe=O intermediate, perhaps yielding a carbon radical and a complex bearing an iron–hydroxy (OH) group, before the formation of a C–OH bond. However, the details of this process are still a source of controversy.

Could a synthetic version of the enzyme reaction be produced?

It is difficult to develop simple artificial

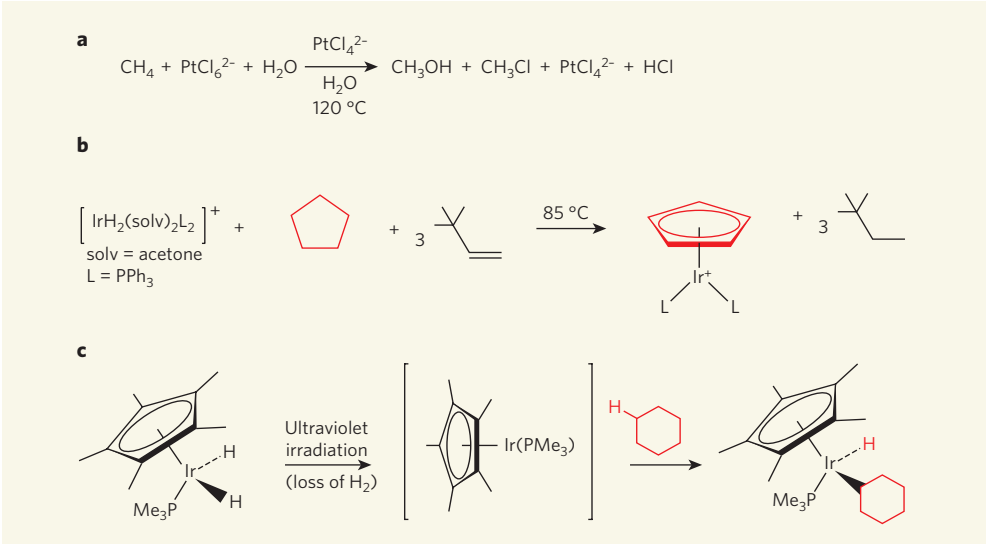
Box 1 | The birth of C–H activation

The origins of the field lie in the 1960s and 1970s, when C–H activation was spotted in saturated molecules in intramolecular reactions (where the metal and the C–H bond are in the same molecule). The first activation reactions of C–H bonds in aromatic molecules were also observed in this period. This led the pioneering organometallic chemist Jack Halpern to comment in 1968 that "the development of successful approaches to the activation of carbon–hydrogen bonds, particularly in saturated hydrocarbons, remains to be achieved and presently constitutes one of the most important and challenging problems in this whole field".

At about the same time, several groups observed that hydrogen atoms in saturated hydrocarbons (alkanes) could be replaced with isotopic deuterium atoms using metal complexes. Even more intriguing were studies showing that alkane hydrogens could be directly replaced with other atoms or chemical groups using platinum (Pt) salt complexes (**a** in the figure). Although these reactions were rather inefficient, they converted simple alkanes into more complex molecules that contained new carbon–chlorine or carbon–oxygen bonds. The mechanisms underlying these processes are still not completely understood today.

In the late 1970s, it was found that certain organometallic compounds (made with the metals rhenium and iridium, Ir) could be used to remove hydrogen from alkanes, so yielding unsaturated hydrocarbons (**b** in the figure; P is a phosphorus atom, Ph an aromatic phenyl group). Light-induced reactions of organometallic tungsten complexes with C–H bonds in aromatic compounds and in silanes — silicon-based equivalents of hydrocarbons — were also observed.

But the field really came alive in the early 1980s, when two groups reported that iridium from an organometallic complex could be inserted between the atoms of C–H bonds in simple alkanes — the first direct observations of oxidative addition (**c** in the figure; Me is a methyl group, CH_3). Shortly afterwards, lutetium-based C–H activation processes were discovered. These findings led to an explosion in C–H activation research that continues to expand to this day. **R.G.B.**



versions of enzymes, because the whole enzyme structure, not just the active site, is often required to provide a selective C–H activation reaction. A few synthetic systems have been generated that use multiply bonded metals to react with hydrocarbons, but these involve metals such as zirconium, titanium and tantalum, which aren't typically found in enzymes. The latest efforts aim to exploit ruthenium and iron as catalytic metals, so mimicking biological systems more closely. Several synthetic iron complexes have been made that use oxygen, or other oxidants such as hydrogen peroxide, to convert alkanes into oxygenated products. These reactions might proceed via transient intermediates that bear Fe–O bonds, but it is difficult to say whether such intermediates are the same as those seen in enzymes because, in most cases, they cannot be easily isolated and characterized.

What other mechanistic questions are there?

Perhaps the biggest issue is whether short-lived alkane complexes, in which C–H bonds bind weakly to metals, act as intermediates in C–H activation reactions that occur by oxidative addition. Direct evidence for this is starting to come from infrared flash kinetics studies, which use rapid laser-pulses to probe reaction intermediates via infrared spectroscopy. Several groups have detected intermediates with very short lifetimes that have been assigned alkane-complex structures. Liquefied xenon and krypton were used as the ultimate inert solvents in one such study. This enabled data to be obtained that were unsullied by solvent interference with the C–H activation reactions under investigation.

What other ways are there to study these intermediates?

In a brilliant piece of work, alkane-complex intermediates have been detected and characterized at low temperatures using nuclear magnetic resonance spectroscopy. A couple of complexes have even been isolated as solids and characterized by X-ray diffraction, although in these cases the complexes fall apart in solution. All these mechanistic studies provide information that is essential for designing improved C–H activation reactions.

What else needs to be done?

More reactions must be developed, so that C–H bonds can be activated as required in any molecule. Although some useful catalytic C–H activation reactions now exist, few can be applied to a broad range of compounds, and many of these work best in unsaturated molecules. Furthermore, the temperatures required for these reactions are often inconveniently high, too much catalyst is needed, and the required reagents are often corrosive and/or expensive. Nevertheless, such transformations are routinely used — for example by chemists in the pharmaceutical industry preparing drug candidates — and they will inevitably soon find their way into industrial applications. Effective methods for the catalytic activation of C–H bonds in alkanes are still rare, but there are encouraging signs of progress, such as the rhodium and platinum chemistry described earlier. These reactions, which are unprecedented using classical chemistry, demonstrate the power of C–H activation to unlock the potential of inert hydrocarbons, and tantalize us with the promise of discoveries yet to come. ■

Robert G. Bergman is in the Department of Chemistry, University of California, Berkeley, Berkeley, California 94270, USA.
e-mail: rgberg@socrates.berkeley.edu

FURTHER READING
Arndtsen, B. A., Bergman, R. G., Mobley, T. A. & Petersen, T. H. *Acc. Chem. Res.* **28,** 154–162 (1995).
Labinger, J. A. & Bercaw, J. E. *Nature* **417,** 507–514 (2002).
Periana, R. A. *et al. J. Mol. Catal. A: Chem.* **220,** 7–25 (2004).
Labinger, J. A. *J. Mol. Catal. A: Chem.* **220,** 27–35 (2004).
Goldberg, K. I. & Goldman, A. S. (eds) *Activation and Functionalization of C-H Bonds* (ACS Symp. Ser. 885, Am. Chem. Soc., Washington DC, 2004).
Tobisu, M. & Chatani, N. *Angew. Chem. Int. Edn* **45,** 1683–1684 (2006).

macroscopic label, as in (*S*)(+)- and (*R*)(−)-CHFClBr in Figure 1. Here, *R* (for *rectus*) and *S* (for *sinister*) are the absolute configurations in the standard nomenclature[3].

Optical rotation and circular dichroism (a small difference in the absorption of right- and left-circularly polarized light) at visible and ultraviolet wavelengths measure natural optical activity in the electronic spectrum of a molecule. It had long been appreciated that extending natural optical activity into the vibrational spectrum could provide more detailed and reliable stereochemical information, because a vibrational spectrum contains many more bands sensitive to the details of the molecular structure[4]. This was finally achieved in the early 1970s, when vibrational optical activity was first observed in small chiral molecules in fluid media using two complementary techniques: a circular-polarization dependence of vibrational Raman scattering of visible laser light[6], and circular dichroism of infrared radiation[7]. These are now known as Raman optical activity (ROA) and vibrational circular dichroism (VCD), respectively.

Haesler *et al.*[1] tested the limits of a new ROA instrument developed in their laboratory by applying it to a chiral molecule for which it would be very difficult, if not impossible, to determine the absolute configuration using any other existing physical technique. They chose (*R*)-[^{2}H$_1$,^{2}H$_2$,^{2}H$_3$]-neopentane, which is C(CH$_3$)$_4$ rendered chiral through isotopic substitution — deuteration — of three of the four methyl groups (Fig. 2). In other words, chirality was induced by replacing ordinary hydrogen (^{1}H) in three of the four CH$_3$ groups with deuterium (^{2}H, or D), producing one each of CH$_3$, CH$_2$D, CHD$_2$ and CD$_3$.

Haesler and colleagues' synthesis[1] of (*R*)-[^{2}H$_1$,^{2}H$_2$,^{2}H$_3$]-neopentane with high enantiomeric purity was itself an achievement, and was attempted only after calculations[8] suggested that the ROA might be measurable. It is a gas at room temperature, but was collected as a liquid for ROA measurements by distilling into a cooled capillary tube that was subsequently sealed. Three distinct orientations of each of the CH$_2$D and CHD$_2$ groups are possible, generating nine distinct rotational 'conformers' that all contribute to chiroptical observable properties and tend to cancel.

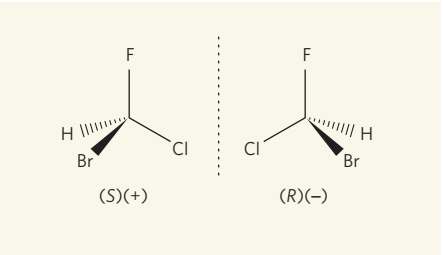


Figure 1 | Chiral archetype. Lord Kelvin's classic definition of chirality[2] emphasizes the non-superposability of a chiral object and its mirror image, exemplified here by the enantiomers of bromochlorofluoromethane, CHFClBr.

The ROA turned out to be just detectable at the limit of sensitivity of the authors' new instrument, which measures tiny circularly polarized components in the vibrational Raman bands down to a few parts in 10^{-5} of the total band intensity (Fig. 2). This instrument separates the right- and left-circularly polarized components of the backscattered light and collects them into the ends of two fibre-optic strands. This allows separate Raman spectra for the right- and left-circularly polarized components of the scattered light to be dispersed simultaneously, one above the other, onto a multichannel detector. Subtraction then provides the required ROA spectrum. Because the right- and left-circularly polarized Raman spectra are measured during the same acquisition period, the 'flicker noise' arising from dust particles, density fluctuations, fluctuations in laser power and so on cancels out, resulting in greatly superior signal-to-noise characteristics compared with earlier designs of the instrument. Other novel features provide a high degree of suppression of the spurious signals that can plague this type of delicate polarization measurement.

Haesler *et al.*[1] simulated the ROA spectrum of the chiral neopentane using *ab initio* quantum-chemical protocols[8]. They summed the partially cancelling ROA spectra computed for each of the nine possible rotational conformers present at thermal equilibrium. The resulting theoretical spectrum was close enough to the observed spectrum with respect to the signs and magnitudes of the ROA bands for the *R* absolute configuration to be assigned unequivocally to their neopentane sample. This highlights the power of vibrational optical activity generally — VCD (ref. 9) as well as ROA — for determining absolute configuration: as there are numerous bands, a calculation that reproduces the signs of most, if not all, of them correctly will provide an absolute configuration on which one might stake one's life.

As well as probing the most delicate sources of chirality in small molecules, ROA provides valuable information about very large chiral molecular structures, such as proteins, sugars and nucleic acids[10]. Even intact viruses are accessible to ROA measurements, and from these, for instance, details about the folds of coat proteins and the nucleic-acid structure may be deduced. The power of chiroptical spectroscopic techniques for biomolecular applications derives from their ability to cut through the complexity of conventional spectra (which are 'blind' to chirality), to reveal three-dimensional information about the most rigid, twisted chiral parts of the structure. These are usually found within the backbone in large biomolecules, and generate the largest chiroptical signals. The decisive advances in ROA instrumentation reported by Haesler *et al.*[1], backed up by powerful theoretical simulation techniques, should facilitate widespread exploitation of vibrational optical activity in chemical and biomolecular science. ■

Laurence D. Barron is in the Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK.
e-mail: laurence@chem.gla.ac.uk

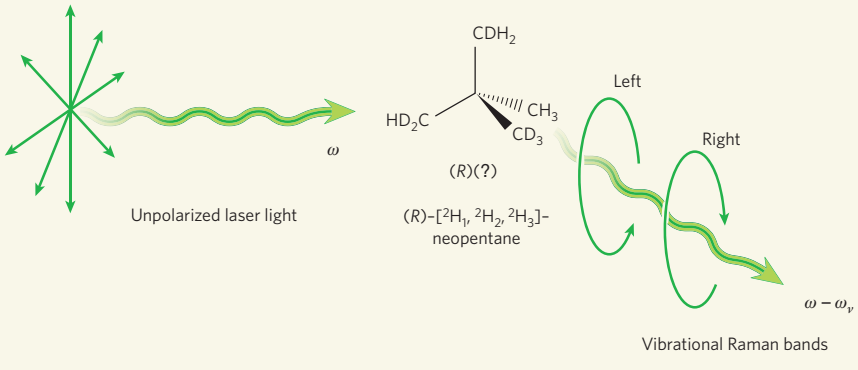


Figure 2 | Determination of the absolute configuration of (*R*)-[^{2}H$_1$,^{2}H$_2$,^{2}H$_3$]-neopentane. Unpolarized laser light incident on the deuterated molecule produces tiny but measurable right- or left-circularly polarized components in the bands of the resulting vibrational Raman spectrum. ω is the angular frequency of the incident visible laser beam and $\omega - \omega_v$ that of one of the many vibrational Raman bands. By comparing these measurements with the expectations of *ab initio* simulations, Haesler *et al.*[1] could assign the *R* absolute configuration to their sample. No (+) or (−) sign is associated with the *R* absolute configuration because there was insufficient material to measure the corresponding optical rotation.

1. Haesler, J., Schindelholz, I., Riguet, E., Bochet, C. G. & Hug, W. *Nature* **446,** 526–529 (2007).
2. Lord Kelvin *Baltimore Lectures* (Clay, London, 1904).
3. Eliel, E. L. & Wilen, S. H. *Stereochemistry of Organic Compounds* (Wiley, New York, 1994).
4. Barron, L. D. *Molecular Light Scattering and Optical Activity* 2nd edn (Cambridge Univ. Press, 2004).
5. Bijvoet, J. M. *et al. Nature* **168,** 271–272 (1951).
6. Barron, L. D., Bogaard, M. P. & Buckingham, A. D. *J. Am. Chem. Soc.* **95,** 603–605 (1973).
7. Holzwarth, G. *et al. J. Am. Chem. Soc.* **96,** 251–252 (1974).
8. Hug, W. & Haesler, J. *Int. J. Quantum Chem.* **104,** 695–715 (2005).
9. Stephens, P. J. & Devlin, F. J. *Chirality* **12,** 172–179 (2000).
10. Barron, L. D. *Curr. Opin. Struct. Biol.* **16,** 638–643 (2006).

Correction
In the News & Views Q&A article "Organometallic chemistry: C–H activation" by Robert G. Bergman (*Nature* **446,** 391–393; 2007), the e-mail address given for the author was incorrect. The correct address is rbergman@berkeley.edu

REVIEWS

Relativistic effects in homogeneous gold catalysis

David J. Gorin[1] & F. Dean Toste[1]

Transition-metal catalysts containing gold present new opportunities for chemical synthesis, and it is therefore not surprising that these complexes are beginning to capture the attention of the chemical community. Cationic phosphine–gold(I) complexes are especially versatile and selective catalysts for a growing number of synthetic transformations. The reactivity of these species can be understood in the context of theoretical studies on gold; relativistic effects are especially helpful in rationalizing the reaction manifolds available to gold catalysts. This Review draws on experimental and computational data to present our current understanding of homogeneous gold catalysis, focusing on previously unexplored reactivity and its application to the development of new methodology.

Interest in the catalytic chemistry of Au(I) complexes has undergone a marked increase. Although traditionally undeveloped as homogeneous catalysts, such species have recently been employed in a plethora of organic transformations. The strong Lewis acidity of cationic Au(I), coupled with its potential to stabilize cationic reaction intermediates, imparts unique reactivity to such catalysts, which has been exploited in the development of new synthetic methods. As the reactivity of Au(I) complexes continues to be uncovered experimentally, the theoretical underpinnings of these new observations warrant consideration. Relativistic effects provide a theoretical framework for rationalizing the observed reactivity—the contracted 6*s* orbital and expanded 5*d* orbitals account for the attributes of Au(I) catalysts. Thus, integrating theoretical and synthetic studies of Au(I) provides a deeper understanding of the fundamental properties of Au and illuminates further avenues for study.

Theoretical chemistry of gold: practical considerations

The relation between structure and function has become a model for chemistry and biochemistry, and thus theoretical studies of the electronic structure of Au provide insight a priori into how it might function as a catalyst and what reactive pathways might be accessible. The theoretical chemistry of Au has recently been well reviewed, and only a few salient points are highlighted here[1,2]. The most prominent characteristics of the electronic structure of Au are the consequence of strong relativistic effects (see Box 1)[3,4]. In Au, the relativistic contraction of the 6*s* orbital results in greatly strengthened Au–L bonds, where L is the ligand (Fig. 1)[5]. Beyond the observed large Au–L bond strengths, further experimental corroboration of this distortion from the otherwise expected electronic structure lies in the phenomenon of 'aurophilicity'[6], the tendency for Au–Au interactions to be stabilizing on the order of hydrogen bonds, and also in the large first ionization potential observed for Au (9.22 eV versus 7.57 eV for Ag).

In contrast with the prevalence of tricoordinate and tetracoordinate Cu(I) and Ag(I) complexes, Au(I) predominantly adopts a linear, bicoordinate geometry[7,8]. The practical consequence of the limited coordination motifs typically observed in Au(I) chemistry is the general need to abstract a ligand from neutral bicoordinate Au(I) species so as to induce catalytic reactivity. Additionally, organoaurate(I) species are not particularly nucleophilic relative to the corresponding copper complexes. Theoretical studies indicate that the Au 5*d* electrons are held with greater energy than the Cu 3*d* electrons due to decreased electron/electron repulsion in the diffuse 5*d* orbitals, resulting in less nucleophilic metal species that do not tend to undergo oxidative addition[9]. Computational and experimental studies on reductive elimination from $LR_3Au(III)$ complexes show such a process to be relatively disfavoured as well[10,11]. These observations are consistent with the broadly observed reactivity of Au(I) and Au(III) complexes, which do not readily cycle between oxidation states, although such reactivity is accessible[12]. Au(I) species are therefore generally tolerant of oxygen, indicating that reactions catalysed by Au might be run without precautions to exclude air. Beyond this practical benefit, the apparent redox stability of Au(I) complexes under ambient conditions allows the development of new modes of reactivity by precluding the traditional oxidative addition/reductive elimination cycles so prevalent in late transition-metal catalysis.

An alkynophilic Lewis acid?

Many of the investigations into the catalytic reactivity of Au exploit the propensity of both Au(III) and cationic Au(I) complexes to activate alkynes towards nucleophilic addition. Research in this area has been extensively reviewed[13,14], and only selected highlights will be presented here. Early investigations into the potential of Au to catalyse synthetic transformations used primarily Au(III) halides and it was found that a wide array of nucleophiles could be added to alkynes in an intramolecular[15–18] or intermolecular[19] fashion. These initial studies aroused the interest of researchers pursuing green chemistry, who have since exploited the apparent insensitivity of Au to aqueous conditions to develop a series of reactions using water or alcohols as solvent[20,21].

In 1998, a seminal contribution by Teles *et al.* demonstrated the utility of cationic phosphineAu(I) species in catalysing the hydration of alkynes[22]. Subsequently, and especially since 2003, Au(I) species have featured more prominently in Au catalysis. R_3PAuX (X = trifluoromethanesulphonate ($^-$OTf) or other weakly coordinating counterions), formed *in situ* by the abstraction of Cl^- from R_3PAuCl by AgX or by the protonolysis of R_3PAuCH_3 with acid, were shown to be superb catalysts for a series of C–C bond-forming reactions, including Conia-ene[23,24] and hydroarylation[25–27] reactions, as well as several carbon–heteroatom bond-forming reactions[28–30]

[1]Department of Chemistry, University of California at Berkeley, Berkeley, California 94720, USA.

Box 1 | Relativistic effects

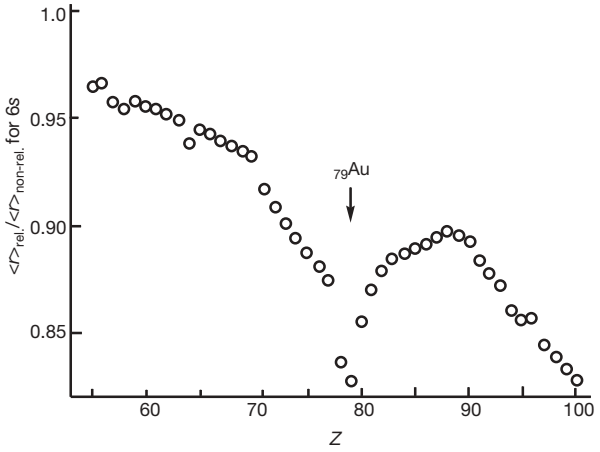
The theoretical underpinning for relativistic effects arises from the confluence of quantum mechanics and special relativity[3,96]. Schrödinger's equation, unveiled in 1926, correctly predicts the atomic orbital energy levels for hydrogen but is unable to account for the fine structure of the hydrogen atomic spectrum, in which the bands are split. Although accounting for spin as a perturbation of the Schrödinger equation corrects for this, in systems in which electrons move at speeds approaching the speed of light (c) a more general relativistic consideration is required. In 1928 Dirac developed a new equation taking special relativity into account, thereby permitting solutions to systems in which electrons move at significant velocities. The term 'relativistic effects' therefore refers to any phenomenon resulting from the need to consider velocity as significant relative to the speed of light.

One basic consequence of the special theory of relativity is that mass increases towards infinity as a body's velocity approaches c, which can be expressed mathematically as $m = m_0/\sqrt{[1-(v/c)^2]}$, where m is the corrected mass, m_0 is non-relativistic (rest) mass, and v is velocity. For a given atom, the average radial velocity of the 1s electrons is $V_r = Z$, where Z is the atomic number. The expression v/c can therefore be calculated as $Z/137$ (137 atomic units (a.u.) = c). For example, in Hg, $Z = 80$ and v/c for the 1s electrons is $80/137 = 0.58$; that is, the 1s electrons have a radial velocity that is 58% of c.

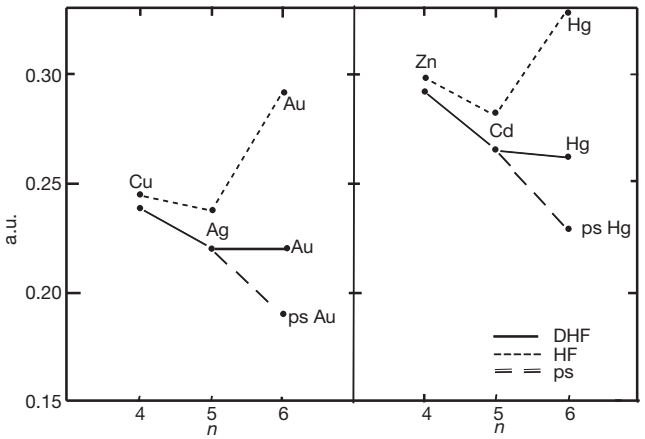
There are three major phenomena that result from relativistic effects[4,97]: the first can easily be rationalized by considering the equations above. In non-relativistic calculations, $c = \infty$ and v/c therefore approaches 0, so no mass correction need be applied to the particles under consideration. In situations in which c is considered to be 137 a.u., the mass of an electron will increase considerably. Because the Bohr radius of an electron orbiting a nucleus is inversely proportional to the mass of the electron, this increase in mass corresponds to a decrease in radius. This relativistic contraction of the 1s orbital also applies to all other s and p orbitals. Thus, the electrons are closer to the nucleus and have greater ionization energies. Practically, this contraction is only significant for elements in which the 4f and 5d orbitals are filled (Box 1 Figure 1).

The second manifestation of relativistic effects is indirect; electrons occupying the d and f orbitals are better shielded by the electrons in the contracted s and p orbitals and therefore see a weaker nuclear attraction. The third effect of a relativistic treatment is spin–orbit coupling, which accounts for the fine splitting in the hydrogen atomic spectrum noted above.

Relativistic effects are crucial to understanding the electronic structure of heavy elements; consideration of other phenomena, such as the lanthanide contraction, are insufficient (Box 1 Figure 2)[98,99]. Among the experimental observations conventionally explained by relativistic effects is the colour of Au. The golden colour is due to excitation of the 5d electrons to the Fermi level, which occurs with a bandgap of 2.38 eV; blue visible light is therefore absorbed. In silver, by contrast, the bandgap is much larger and no visible light is absorbed. The smaller bandgap in Au is due to the relativistic contraction of the 6s and 6p orbitals and the expansion of the 5d orbitals.



Box 1 Figure 1 | Calculated relativistic contraction of the 6s orbital. The relativistic and non-relativistic 6s orbital radii were determined computationally[100]. Notably, Pt, Au and Hg are markedly influenced. (Reprinted from ref. 4, with permission from the American Chemical Society.)



Box 1 Figure 2 | Calculated ionization potentials for the Group 11 and Group 12 transition metals. Three different methods are compared: Dirac–Hartree–Fock (DHF) (relativistic), Hartree–Fock (HF) (non-relativistic), and pseudopotential (ps) (without regard for f electrons). The difference between the HF and ps calculations results from accounting for the lanthanide contraction; the difference between DHF and HF results from accounting for relativistic effects. The experimental ionization potentials closely match those given by DHF. (Reprinted from ref. 99, with permission from Elsevier.)

(Fig. 2). Gold has also been used with some success in additions of nucleophiles to allenes[31–33], alkenes[34,35] and dienes[36,37], although whether these reactions proceed by means of activation of the π-system by Au remains an area of active investigation, especially for additions to unactivated alkenes[38,39].

Cationic Au(I) species are superior Lewis acids compared with other Group 11 metals for many transformations, and intuitively it seems that relativistic contraction of the valence s or p orbitals of Au should be responsible, because they should correspond to a relatively low-lying lowest unoccupied molecular orbital (LUMO) and therefore strong Lewis acidity. This conclusion can also be derived from considering the high electronegativity of Au (2.4, compared with 1.9 for Ag): strong Lewis acidity generally correlates with electronegativity (provided that a coordination site is available for accepting an electron pair). It is crucial to realize that the electronegativity of Au is a result of the relativistic contraction of the valence 6s and 6p orbitals; thus, consideration of relativistic effects does not merely corroborate an analysis based on electronegativity but actually provides a theoretical underpinning to understand it.

From a computational standpoint, natural bond orbital (NBO) population analyses of the Au atomic orbital occupancies for cationic Au(I)–phosphine complexes indicated that for $AuPH_3^+$ the 6s occupancy was 0.438 electrons, whereas for the corresponding complex $AgPH_3^+$ the 5s occupancy was only 0.156 (refs 8, 40). This suggested that the Au–phosphine bond was far more covalent than the Ag–phosphine bond, an unsurprising notion considering the comparison of AuH and AgH presented in Fig. 1, and complemented by the decreased charge density calculated on Au in comparison with that on Ag (0.66 versus 0.86).

As models for complexation of a second Lewis basic ligand, $Au(PH_3)_2^+$ and $Ag(PH_3)_2^+$ were considered. The Au 6s occupancy increased twofold, to 0.796, whereas the Ag 5s occupancy was 0.449

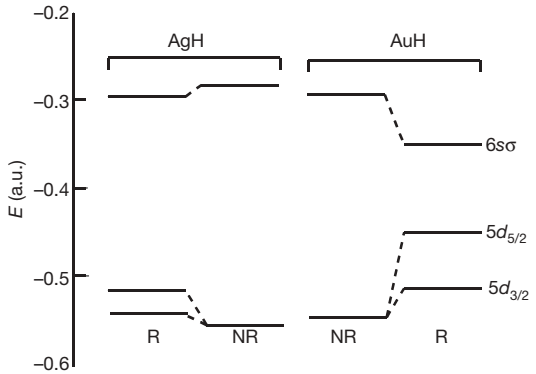


Figure 1 | Comparison of AuH and AgH bond energies. Calculated (Hartree–Fock) molecular orbital energies for AgH and AuH are shown with (R) and without (NR) consideration of relativistic effects[5]. In the non-relativistic calculation, the two molecules had similar bond energies. In the relativistic calculation, the 6*s* contraction resulted in a far stronger AuH σ-bond. (Reprinted from ref. 4, with permission from the American Chemical Society.)

(refs 8, 40). This suggests rehybridization of the phosphineAu(I) molecular orbitals to resemble a three-centre four-electron bond that greatly increases the occupancy of the *s* orbital. The other orbitals that might accept electrons, the 6*p*, remained largely unoccupied (0.011 electrons). Although simplistic, this consideration, if generalized, can rationalize the superior Lewis acidity of Au, because the low-lying *s* orbital should be responsible for binding in phosphineAuL$^+$. Additionally, because Au(I)$^+$ is a large, diffuse cation that shares positive charge with the phosphine ligand, one might expect orbital rather than charge interactions to dominate in binding a second ligand. Thus, phosphineAu(I)$^+$ may be considered a 'soft' Lewis acid, preferentially activating 'soft' electrophiles, such as π-systems.

Although the 'soft' Lewis acidity of Au(I) can be rationalized, further consideration is needed to understand the superior reactivity of Au–alkyne complexes towards nucleophilic addition. Studies on Au$^+$–ethylene and Au$^+$–ethyne bonding indicate ~10 kcal mol^{-1} greater stabilization for the ethylene complex over the ethyne complex[41,42]. Because Au(I) apparently does not selectively complex alkynes over other π-systems, the observed reactivity may be due to discrimination by the nucleophile in selecting between Au(I)-activated electrophiles. Such selectivity can be understood by considering the relative energy levels of ethyne and ethylene LUMOs: alkynes have intrinsically lower highest occupied molecular orbitals and LUMOs than the corresponding alkenes (by ~0.5 eV), and are therefore less nucleophilic and more electrophilic[43]. It can therefore be expected that an Au–alkyne complex should have a lower LUMO for the addition of a nucleophile than an analogous Au–alkene complex; this is potentially the source of the 'alkynophilicity' observed in Au(I)$^+$-catalysed reactions.

The mechanism of Au(I)-mediated activation of an alkyne towards addition by methanol was originally investigated by the Teles group in calculations that predicted a *cis*-oxyauration of the alkyne[22,44]. In contrast, labelling experiments by the Toste group in the Conia-ene carbocyclization[23] and Hashmi *et al.* in the cycloisomerization of propargyl amides[45] are consistent with *anti*-addition of Au and a nucleophile across an alkyne.

Initial explorations demonstrated some effect of oxidation state on Au catalysis: the regioselectivity was reversed for the intermolecular hydroarylation of alkynes with arenes in the presence of Au(I) as opposed to Au(III)[46]. More telling work by Gevorgyan on the Au(I)- and Au(III)-catalysed cyclization of halogenated allenones to form halofurans offers insight into the mechanistic pathways available in reactions catalysed by Au (Fig. 3)[47]. In this case, catalysts in each oxidation state led to divergent products. With Au(I) a product consistent with activation of the allene was observed, whereas with Au(III) the product formed was consistent with activation of the ketone moiety. Unrelated calculations by Straub support this notion of functional group discrimination, indicating that $AuCl_3$ exhibits a thermodynamic preference for aldehyde coordination over alkyne coordination by 21.3 kJ mol^{-1}, while also being capable of catalysing reactions through alkyne-activation pathways[48].

Especially for Au(I), ligand effects can influence reaction outcome; Au(I)-phosphine species are well studied[8,40], offering a tunable ligand set for optimizing catalyst reactivity. In the Au(I)-catalysed ring expansion of propargylcyclopropanols, [(4-CF_3-C_6H_4)$_3$P]AuCl was shown to be superior to other Au-phosphines[49]. Buchwald-type ligands were most effective in the intramolecular hydroamination of allenes[50], whereas N-heterocyclic carbene (NHC) ligands were optimal for an indene synthesis[51]. Additionally, the trimeric Au

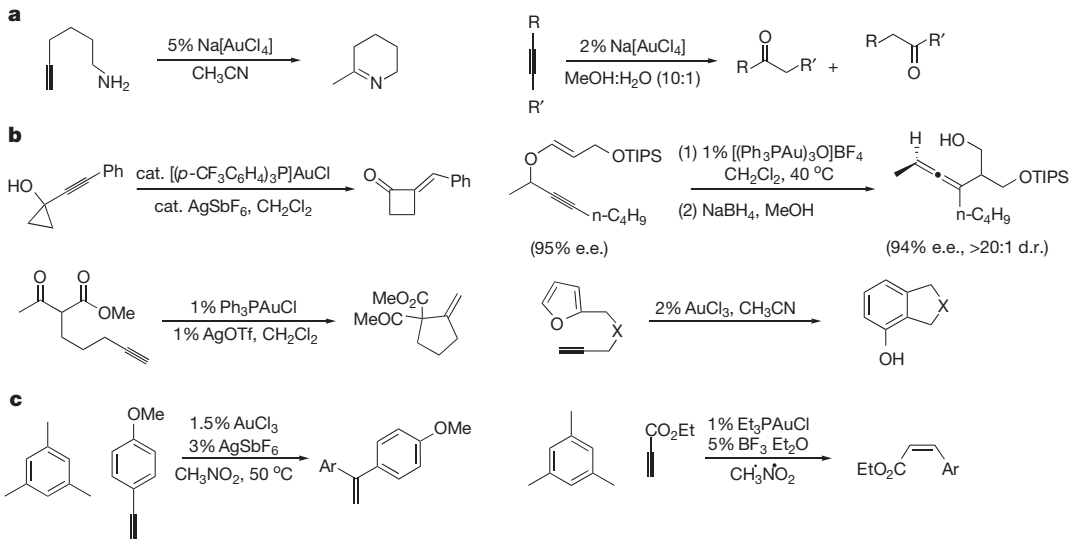


Figure 2 | Catalytic activation of alkynes. Au(III) and Au(I)$^+$ are excellent Lewis acids capable of catalysing the addition of a wide variety of nucleophiles to alkynes. **a**, Representative examples featuring heteroatom–carbon bond formation include hydroamination[15] and hydration[19] reactions. **b**, Activated carbon nucleophiles have been extensively investigated: the illustrated ring expansion[49], Conia-ene[23], propargyl Claisen[52] and cycloaddition/fragmentation[17] carbon–carbon bond-forming reactions are among those reported. d.r., diastereomeric ratio; e.e., enantiomeric excess; TIPS, tri-(isopropyl)silyl). **c**, Hydroarylation has been studied with Au(I) and Au(III): different reactivity is observed depending on the oxidation state of Au (ref. 25).

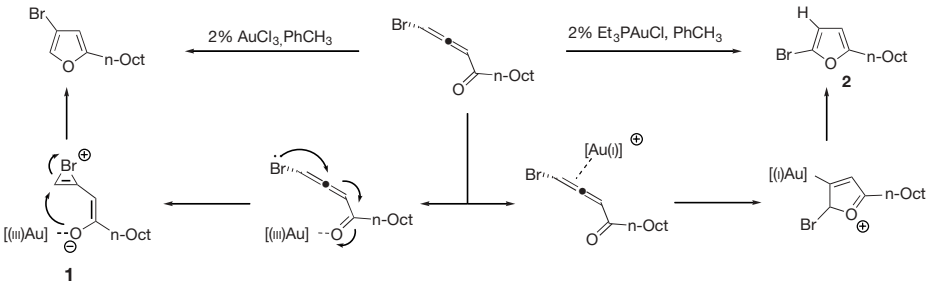


Figure 3 | Oxidation state determines product. Choice of Au source dictated the product distribution obtained from the cyclization of halogenated allenones[47]. The observed results are consistent with mechanisms in which Au(III) activated the ketone, leading to isomerization through the bromonium intermediate **1**, whereas Au(I) coordinated the allene, leading to product **2** in which no halogen migration occurred. Oct, $(CH_2)_7CH_3$.

complex $(PPh_3Au)_3OBF_4$ was superior to monomeric Au phosphines as a catalyst for the propargyl Claisen rearrangement[52]. Although some initial studies to improve the stability of Au(III) catalysts through the use of ligands have been reported[53,54], most efforts have focused on Au(I).

Relativistic effects and backbonding

Although Au(I) is a versatile catalyst for activating alkynes towards nucleophilic addition, a question arose: are Au complexes only capable of mediating reactions by functioning as Lewis acids? After all, innumerable possible Lewis and Brönsted acids might similarly activate alkynes. One particularly intriguing calculation suggesting a coordination mode other than simple Lewis acid–base interaction was performed by Irikura & Goddard, who determined that the $AuCH_2^+$ fragment **3** should feature multiple bond character, through σ-complexation of singlet CH_2 and backbonding from Au to methylene (Fig. 4)[55]. Schwarz undertook similar calculations in evaluating the extent of relativistic effects on cationic late-metal carbene complexes; it was found that although $AuCH_2^+$ possessed the lowest bond-dissociation energy among the metal carbenes examined, more than 70% of the total bond energy for Au was a contribution from relativistic effects[56]. More recently, Barysz & Pyykkö proposed that AuC^+ **4** should have some triple-bond character[57], and mass spectroscopic evidence for such a species has been accumulated[58].

These calculations indicating significant backbonding from Au(I) into vacant *p* orbitals contrast the absence of backbonding observed in Au–carbonyl and Au–alkyne complexes. Both $[Au(CO)]^+$ and $[Au(CO)_2]^+$ have been characterized spectroscopically[59]. The ν_{CO} (2,194 and 2,208 cm^{-1}) are actually greater than that of free CO, suggesting σ-only complexation of CO to Au^+ without backbonding from Au into the ligand π* orbital. Calculations on Au(I)–ethylene and Au(I)–ethyne complexes indicate that backbonding accounts for a far smaller fraction of the bonding energy than in the analogous Cu complexes[41,42]. Thus, $Au(I)^+$ cannot be considered to participate significantly in Chatt–Dewar–Duncanson-type bonding[60,61] because antibonding orbitals are apparently too high in energy for meaningful backbonding to occur, whereas lower-energy non-bonding *p* orbitals may be more suitable for overlap with the filled Au 5*d* orbitals. This may be relevant to the Lewis acidity of Au(I); the lack of backbonding from Au(I) into π-ligands could render the ligands more electron deficient, contributing to the ease of nucleophilic addition.

Synthetically, limited progress has been made in isolating Au complexes with carbenoid character. Gold–NHC complexes are known, although there is no structural evidence for multiple-bond character in the metal–ligand bond[62]. Transmetalation of Fischer carbenes from Group VI metals to Au has been accomplished, and the resulting metal-complexed vinyl Au species **5** was proposed to have significant Au carbene character on the basis of nuclear magnetic resonance spectroscopy, X-ray crystallography, and calculations[63]. This experiment corroborates the theoretical studies of Au(I)-carbenoid species mentioned earlier and suggests the potential of Au to engage in backbonding into the LUMO of σ-bonded organic cations.

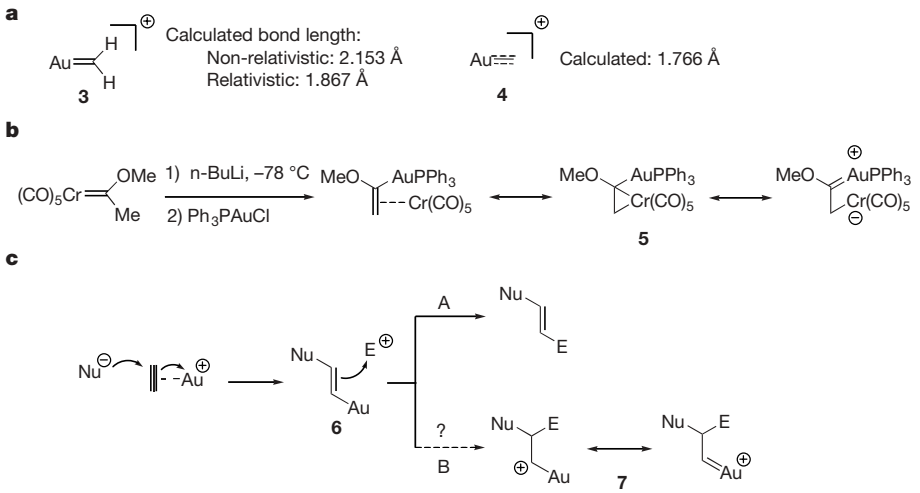


Figure 4 | Proposed multiply-bonded Au–C structures. **a**, Au-alkylidene (**3**) and Au-alkylidyne (**4**) structures have been calculated[56,57], and relativistic effects are posited to be largely responsible for the predicted multiple-bond character. **b**, Experimental data consistent with carbene character in Au species **5** was accumulated in the transmetalation of a chromium Fischer carbene to Au(I)[63]. **c**, Practically, these studies imply that the Au 5*d* electrons backbond into conjugated empty carbon *p* orbitals, and thus Au complexes may be viable for stabilizing cationic intermediates, such as **7**, in catalytic reactions. Nu, nucleophile.

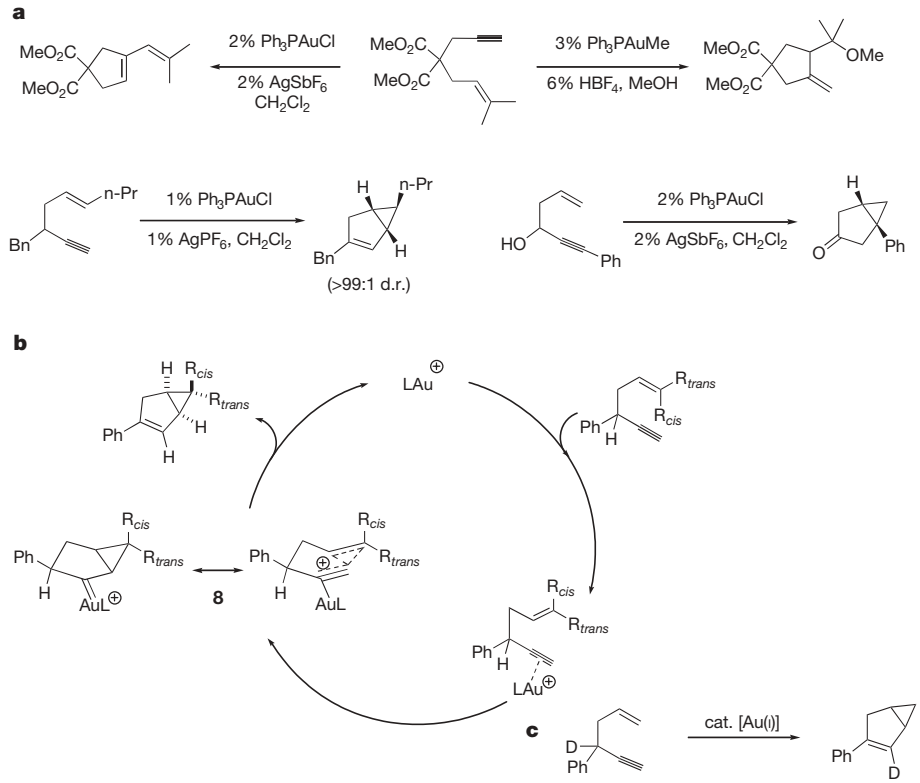


Figure 5 | Enyne cycloisomerization. a, Au(I)-catalysed cycloisomerizations of 1,5- and 1,6-enynes are proposed to proceed through carbocationic intermediates stabilized by Au[66–69]. **b**, The proposed catalytic cycle for the cycloisomerization of 1,5-enynes posits resonance-stabilized intermediate **8**. **c**, Although deuterium labelling provides evidence for the presence of intermediate **8**, the extent of carbene as opposed to carbocationic character is difficult to judge.

In all of the synthetic applications of Au(I) catalysts discussed so far, the proposed vinyl-Au intermediate **6** formed from the initial addition of a nucleophile to an alkyne breaks down by means of protodemetalation (Fig. 4, path A); recently, such intermediates have also been trapped by other electrophiles in an intramolecular fashion[31,64,65]. On the basis of the theoretical data presented above, an alternative manifold may be envisaged (Fig. 4, path B): $Au(I)^+$ would first activate an alkyne towards nucleophilic addition as before; Au could then facilitate trapping of an electrophile by backbonding from the relativistically expanded 5*d* orbitals into the developing conjugated cation **7**.

The first experimental evidence that such a hypothesis might accurately reflect the reactivity of Au came nearly simultaneously in reports by the Echavarren group[66,67], the Fürstner group[68] and the Toste group[69], all dealing with the Au(I)-catalysed cycloisomerization of enynes (Fig. 5). These cycloisomerizations proceed through cationic intermediates, and it was proposed that Au might lower the barrier to accessing these by backbonding to delocalize the positive charge by the formation of carbenoid species. Enynol cyclisomerizations were subsequently performed with Au(III) sources, and analogous intermediates were proposed[70].

The mechanistic hypotheses put forth for the Au-catalysed cycloisomerization reactions are precedented by those posited for similar enyne isomerizations with Pt(II) salts: Fürstner *et al.* have previously discussed a carbene–cation continuum in describing reaction intermediates[71]. The extent of carbenoid as opposed to simple cationic character at the carbon adjacent to the metal is unclear and depends on the catalyst[72]. Divergent product formation in the $InCl_3$- and AuCl-catalysed syntheses of halophenanthrenes suggests that Au(I) species may have significant carbene character whereas other catalysts for alkyne activation, such as In(III)[73] and Ga(III)[74], show more traditional Lewis acidic reactivity (Fig. 6)[75].

Pt(II) shows similar reactivity to Au(I) in many cases[72,76,77], an unsurprising observation given that relativistic effects also heavily influence Pt, and so similarly stabilized cationic intermediates may be accessed. The crucial advantage of Au(I) catalysis is that electrophilic activation of alkynes with cationic Au(I) proceeds even in the presence of strongly donating ligands, such as phosphines, whereas the activation of alkynes by Pt(II) generally proceeds with simple salts or in the presence of CO (ref. 78). Although phosphinePt(II) precatalysts have recently been reported to activate alkynes, the role of the ligand remains unclarified[79,80]. The use of phosphineAu(I) complexes, in contrast, demonstrably permits the tuning of reactivity and, as initially demonstrated in a pioneering study by Hayashi and Ito, stereoselectivity[81].

Although the report by Hayashi and Ito on isocyanate aldol chemistry provided a proof of principle for enantioselective Au

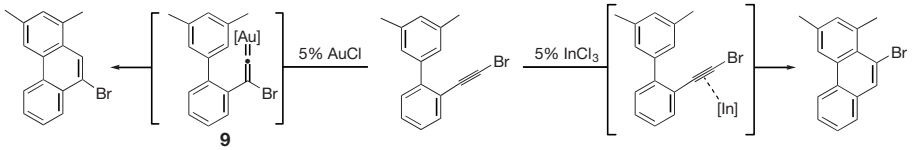


Figure 6 | Comparison of reactivities of Au(I) and In(III). Evidence for Au-carbene character was accumulated in the metal-catalysed synthesis of halophenanthrenes: halogen migration in the reaction with AuCl is consistent with the formation of alkylidene intermediate **9**, whereas catalysis with $InCl_3$ resulted in the product arising from Lewis acidic activation of the alkyne by π-complexation[75].

catalysis, only recently has this been extended to reactions proceeding by means of Au-catalysed activation of alkynes. A recent report from the Echavarren group achieved this, as the isomerization–alkoxylation of 1,6-enynes was investigated in the presence of Au(I)-complexes featuring chiral phosphine ligands[82]. Although only moderate enantiomeric excess (e.e.) could be obtained (most examples had less than 50% e.e., but one had 94% e.e.), these results were crucial in demonstrating that the ligand on Au could create a chiral environment, which was perhaps surprising in view of the preferred linear coordination geometry of Au(I), which necessitates that the source of chirality be far from the centre of induced stereochemistry.

Further illumination in discriminating between carbenoid and cationic intermediates came in the course of studies on the Au(I)-catalysed Rautenstrauch rearrangement of 3-acetoxy-1,4-enynes (Fig. 7)[83]. A carbenoid intermediate in the reaction mechanism seemed reasonable, analogous both to the posited mechanism for the Au-catalysed cycloisomerization of enynes and that proposed for the Pd-catalysed version of the Rautenstrauch[84]. Surprisingly, chirality transfer in the course of the cycloisomerization indicated that formation of a Au carbene did not precede the cyclization, because that would presumably necessitate C–O bond cleavage and concurrent loss of stereochemical information. Therefore, a Nazarov cyclization-like transition state **10** was proposed, in which C–C bond formation occurs in concert with C–O bond cleavage. Subsequent calculations predicted an asynchronous mechanism, in which the C–O bond breaks immediately before bond formation, albeit on a timescale on which stereochemical information is retained in the helical intermediate **11** (ref. 85). Similar reactivity has been observed in the Au(III)-catalysed intramolecular cyclopropanation of 5-acetoxy-1,6-enynes: chirality transfer is again inconsistent with carbene formation[86–88].

Carbenoid reactivity

In light of the chirality transfer observed in the Rautenstrauch reaction, it may be suggested that slowing the rate of trapping of the intermediate generated from propargyl esters in the presence of cationic Au(I) might allow access to intermediates of highly carbenoid character (Fig. 8). One simple way to accomplish this is to render the reaction intermolecular; presumably intramolecular C–O bond cleavage is rapid relative to intermolecular trapping. An Au(I)-catalysed intermolecular cyclopropanation reaction[89], by analogy with the Ru chemistry reported by Ohe and Uemura[90], was therefore developed. In contrast with the Au(I)-catalysed Rautenstrauch rearrangement, stereochemical information from enantioenriched propargyl esters was not conserved. Notably, the cyclopropanation reaction occurred stereospecifically with *cis* and *trans* β-methyl styrene, widely regarded as a benchmark test for a concerted cyclopropanation mechanism and consistent with formation of the carbene intermediate **12**. The intermolecular cyclopropanation is also catalysed by $AuCl_3$, albeit with lower selectivity[90]. Crucially, the use of $(AuCl)_2$(R-DTBM-Segphos) (where DBTM-Segphos is (4,4′-bi-1,3-benzodioxole)-5,5-diylbis(di(4-tert-butyl-3,5-dimethoxyphenyl)-phosphine)) **13** resulted in an enantiomeric excess of up to 94% for the cyclopropanation of a variety of styrene derivatives.

Further methods of probing the backbonding potential of Au are of interest; an intramolecular alternative to the electrophilic trapping

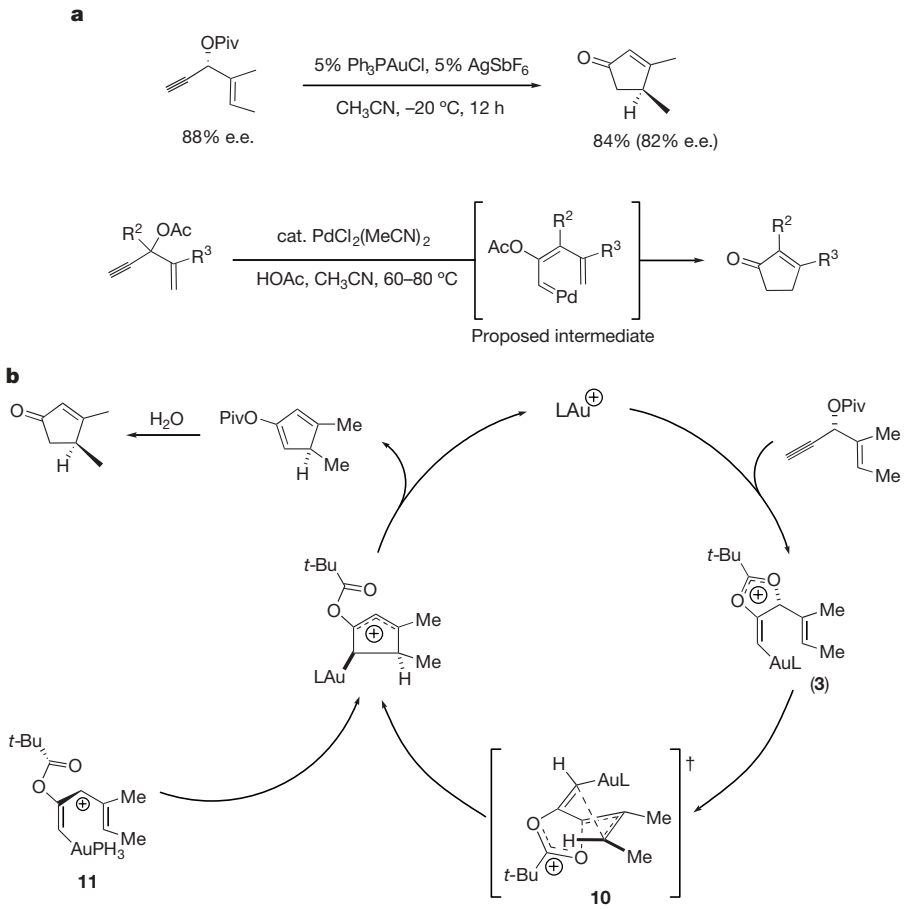


Figure 7 | Mechanism of the Rautenstrauch reaction. a, The Au(I)-catalysed Rautenstrauch rearrangement proceeded with chirality transfer from enantioenriched propargylvinyl esters[83], suggesting a mechanism different from that posited in the Pd-catalysed case because the formation of a carbenoid intermediate should result in racemization[84]. **b**, Two possible mechanisms leading to the conservation of stereochemical information have been proposed: both transition state **10**, indicating a concerted cyclization mechanism, and helically chiral intermediate **11** could account for the observed stereochemistry[85]. Piv, pivaloate.

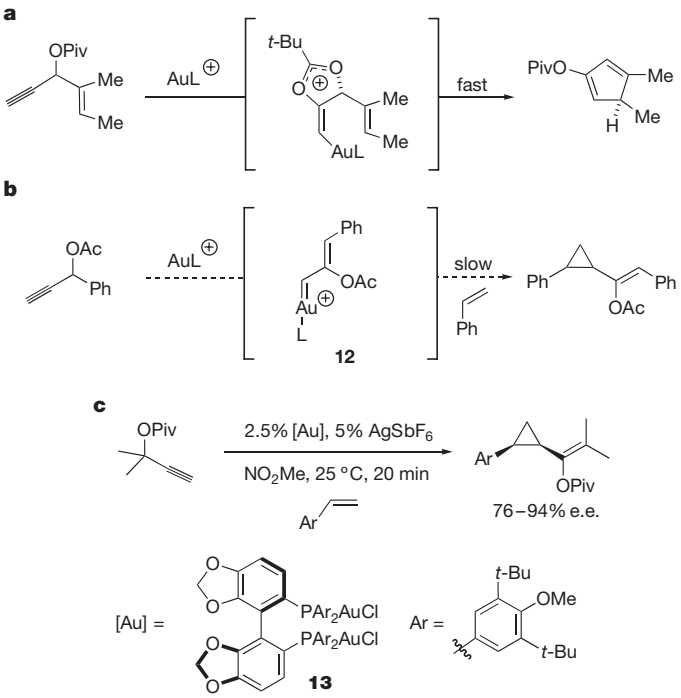


Figure 8 | A proposed carbenoid intermediate. **a**, The rate of intramolecular trapping of the intermediate generated from propargyl esters and $Au(I)^+$ was fast in the Rautenstrauch rearrangement[83]. **b**, It was proposed that intermolecular trapping would be sufficiently slow to permit the formation of carbenoid intermediate **12**. **c**, Thus, an intermolecular cyclopropanation of alkenes with propargyl esters was developed[89]. The result of a chirality transfer experiment contrasted with observations made in the Rautenstrauch cyclization: stereochemical information was not conserved, which is consistent with the proposed intermediacy of **12**. The asymmetric cyclopropanation of styrenes was accomplished with $(AuCl)_2$(R-DTBM-Segphos) **13**. Piv, pivaloate; DTBM, 4-tert-butyl-3,5-dimethoxyphenyl.

presented in Fig. 4 was envisioned. Hypothetically, Au(I) could activate an alkyne towards nucleophilic addition, and then assist in displacing a pendant leaving group (Fig. 9). From this proposal an Au(I)-catalysed intramolecular acetylenic Schmidt reaction was developed[91]. The proposed intermediate **14** in the acetylenic Schmidt reaction could be trapped through ring expansion of carbocycles and intramolecular migration of a siloxy substituent. Judicious choice of ligand was necessary in developing optimal reaction conditions, and $(AuCl)_2(Ph_2PCH_2PPh_2)$ treated with $AgSbF_6$ proved ideal. It is possible that aurophilic interactions between the two Au atoms ligated to the phosphine have a function in stabilizing the catalyst, although the precise benefit derived from the bis-phosphine ligand remains unknown[92].

The reactivity of cationic Au(I) species observed so far is strongly consistent with the stabilization of conjugated cationic intermediates by means of backbonding from Au. Au(III) and Pt(II) catalysts show similar reactivity in many circumstances, although so far phosphineAu(I) complexes are the catalysts of choice for controlling stereoselectivity. The quest for isolated Au complexes with unambiguous carbenoid character continues. Isolation and characterization of catalytic reaction intermediates may be helpful in this regard, alongside efforts to effect carbene transfer to Au[63,93]. More rigorous studies to determine precisely the active catalysts and operative mechanisms in reactions catalysed by Au are clearly needed. In addition, theoretical work directly addressing the superior π-acidity of Au in the presence of alkynes and explaining the electronic structure of proposed intermediates in catalytic cycles will be invaluable in designing new generations of catalysts and reactions.

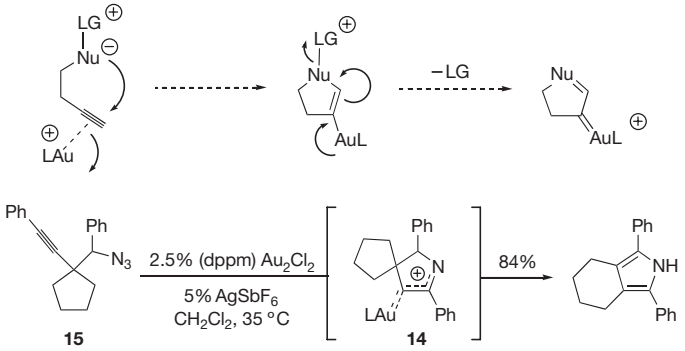


Figure 9 | Intramolecular acetylenic Schmidt reaction. An intramolecular variant on the hypothesis depicted in Fig. 4 was examined by means of the development of a pyrrole synthesis from homopropargyl azides[91]. The cyclization reaction was proposed to proceed through delocalized cation **14**, in which the positive charge is partly delocalized onto Au. Evidence for this organogold intermediate was provided by intramolecular trapping experiments, including ring expansion of carbocycle **15**. dppm, $Ph_2PCH_2PPh_2$; LG, leaving group.

As the upsurge in reviews and reports on the subject testifies, homogeneous Au catalysis invites increasing attention and excitement[94,95]. Gold complexes are superb Lewis acids for the activation of alkynes, a phenomenon best understood in the light of the relativistically contracted 6*s* orbital of Au. Gold carbenoid species are proposed to exist transiently in the course of several catalytic cycles, and this reactivity may be attributed to the relativistically expanded 5*d* orbitals of Au; the 5*d* electrons remain too low in energy to engage in meaningful backbonding to anti-bonding orbitals but are able to delocalize into lower-energy, empty, non-bonding orbitals. The exploitation of Au catalysts in developing new reactivity benefits greatly from a fundamental understanding of the electronic structure of the metal and serves as a model for integrating theoretical and synthetic chemistry in the development of new methodology.

1. Pyykkö, P. Theoretical chemistry of gold. II. *Inorg. Chim. Acta* **358,** 4113–4130 (2005).
2. Pyykkö, P. Theoretical chemistry of gold. *Angew. Chem. Int. Ed.* **43,** 4412–4456 (2004).
3. McKelvey, D. R. Relativistic effects on chemical properties. *J. Chem. Educ.* **60,** 112–116 (1983).
4. Pyykkö, P. & Desclaux, J. P. Relativity and the periodic system of elements. *Acc. Chem. Res.* **12,** 276–281 (1979).
5. Desclaux, J. P. & Pyykkö, P. Dirac–Fock one-center calculations—molecules CuH, AgH and AuH including P-type symmetry functions. *Chem. Phys. Lett.* **39,** 300–303 (1976).
6. Scherbaum, F., Grohmann, A., Huber, B., Krüger, C. & Schmidbaur, H. 'Aurophilicity' as a consequence of relativistic effects: The hexakis(triphenylphosphaneaurio)methane dication $[(Ph_3P)Au]_6C_2^+$. *Angew. Chem. Int. Edn Engl.* **27,** 1544–1546 (1988).
7. Carvajal, M. A., Novoa, J. J. & Alvarez, S. Choice of coordination number in d^{10} complexes of Group 11 metals. *J. Am. Chem. Soc.* **126,** 1465–1477 (2004).
8. Schwerdtfeger, P., Hermann, H. L. & Schmidbaur, H. Stability of the gold(I)–phosphine bond. A comparison with other Group 11 elements. *Inorg. Chem.* **42,** 1334–1342 (2003).
9. Nakanishi, W., Yamanaka, M. & Nakamura, E. Reactivity and stability of organocopper(I), silver(I), and gold(I) ate compounds and their trivalent derivatives. *J. Am. Chem. Soc.* **127,** 1446–1453 (2005).
10. Komiya, S. & Kochi, J. K. Electrophilic cleavage of organogold complexes with acids—mechanism of reductive elimination of dialkyl(aniono)gold(III) species. *J. Am. Chem. Soc.* **98,** 7599–7607 (1976).
11. Komiya, S., Albright, T. A., Hoffmann, R. & Kochi, J. K. Reductive elimination and isomerization of organogold complexes—theoretical studies of trialkylgold species as reactive intermediates. *J. Am. Chem. Soc.* **98,** 7255–7265 (1976).
12. Tamaki, A. & Kochi, J. K. Oxidative addition in coupling of alkylgold(I) with alkyl-halides. *J. Organometall. Chem.* **64,** 411–425 (1974).
13. Hashmi, A. S. K. Homogeneous catalysis by gold. *Gold Bull.* **37,** 51–65 (2004).
14. Hashmi, A. S. K. Homogeneous gold catalysts and alkynes: A successful liaison. *Gold Bull.* **36,** 3–9 (2003).
15. Fukuda, Y., Utimoto, K. & Nozaki, H. Preparation of 2,3,4,5-tetrahydropyridines from 5-alkynylamines under the catalytic action of Au(III). *Heterocycles* **25,** 297–300 (1987).

16. Asao, N., Takahashi, K., Lee, S., Kasahara, T. & Yamamoto, Y. $AuCl_3$-catalyzed benzannulation: Synthesis of naphthyl ketone derivatives from *o*-alkynylbenzaldehydes with alkynes. *J. Am. Chem. Soc.* **124,** 12650–12651 (2002).
17. Hashmi, A. S. K., Frost, T. M. & Bats, J. W. Highly selective gold-catalyzed arene synthesis. *J. Am. Chem. Soc.* **122,** 11553–11554 (2000).
18. Hashmi, A. S. K., Schwarz, L., Choi, J.-H. & Frost, T. M. A new gold-catalyzed C–C bond formation. *Angew. Chem. Int. Ed.* **39,** 2285–2288 (2000).
19. Fukuda, Y. & Utimoto, K. Effective transformation of unactivated alkynes into ketones or acetals by means of Au(III) catalyst. *J. Org. Chem.* **56,** 3729–3731 (1991).
20. Yao, X. & Li, C. J. Water-triggered and gold(I)-catalyzed cascade addition/cyclization of terminal alkynes with *ortho*-alkynylaryl aldehyde. *Org. Lett.* **8,** 1953–1955 (2006).
21. Arcadi, A., Bianchi, G., Di Giuseppe, S. & Marinelli, F. Gold catalysis in the reactions of 1,3-dicarbonyls with nucleophiles. *Green Chem.* **5,** 64–67 (2003).
22. Teles, J. H., Brode, S. & Chabanas, M. Cationic gold(I) complexes: Highly efficient catalysts for the addition of alcohols to alkynes. *Angew. Chem. Int. Ed.* **37,** 1415–1418 (1998).
23. Kennedy-Smith, J. J., Staben, S. T. & Toste, F. D. Gold(I)-catalyzed Conia-ene reaction of beta-ketoesters with alkynes. *J. Am. Chem. Soc.* **126,** 4526–4527 (2004).
24. Staben, S. T., Kennedy-Smith, J. J. & Toste, F. D. Gold-catalyzed 5-endo-dig carbocyclization of acetylenic dicarbonyl compounds. *Angew. Chem. Int. Ed.* **43,** 5350–5352 (2004).
25. Reetz, M. T. & Sommer, K. Gold-catalyzed hydroarylation of alkynes. *Eur. J. Org. Chem.* **2003,** 3485–3496 (2003).
26. Nevado, C. & Echavarren, A. M. Intramolecular hydroarylation of alkynes catalyzed by platinum or gold: Mechanism and endo selectivity. *Chem. Eur. J.* **11,** 3155–3164 (2005).
27. Ferrer, C. & Echavarren, A. M. Gold-catalyzed intramolecular reaction of indoles with alkynes: Facile formation of eight-membered rings and an unexpected allenylation. *Angew. Chem. Int. Ed.* **45,** 1105–1109 (2006).
28. Antoniotti, S., Genin, E., Michelet, W. & Genet, J. P. Highly efficient access to strained bicyclic ketals via gold-catalyzed cycloisomerization of bis-homopropargylic diols. *J. Am. Chem. Soc.* **127,** 9976–9977 (2005).
29. Buzas, A. & Gagosz, F. Gold(I)-catalyzed formation of 4-alkylidene-1,3-dioxolan-2-ones from propargylic tert-butyl carbonates. *Org. Lett.* **8,** 515–518 (2006).
30. Mizushima, E., Hayashi, T. & Tanaka, M. Au(I)-catalyzed highly efficient intermolecular hydroamination of alkynes. *Org. Lett.* **5,** 3349–3352 (2003).
31. Zhang, L. M. Tandem Au-catalyzed 3,3-rearrangement-[2 + 2] cycloadditions of propargylic esters: Expeditious access to highly functionalized 2,3-indoline-fused cyclobutanes. *J. Am. Chem. Soc.* **127,** 16804–16805 (2005).
32. Suhre, M. H., Reif, M. & Kirsch, S. F. Gold(I)-catalyzed synthesis of highly substituted furans. *Org. Lett.* **7,** 3925–3927 (2005).
33. Morita, N. & Krause, N. The first gold-catalyzed C–S bond formation: Cycloisomerization of alpha-thioallenes to 2,5-dihydrothiophenes. *Angew. Chem. Int. Ed.* **45,** 1897–1899 (2006).
34. Zhang, J. L., Yang, C. G. & He, C. Gold(I)-catalyzed intra- and intermolecular hydroamination of unactivated olefins. *J. Am. Chem. Soc.* **128,** 1798–1799 (2006).
35. Yao, X. & Li, C. J. Highly efficient addition of activated methylene compounds to alkenes catalyzed by gold and silver. *J. Am. Chem. Soc.* **126,** 6884–6885 (2004).
36. Brouwer, C. & He, C. Efficient gold-catalyzed hydroamination of 1,3-dienes. *Angew. Chem. Int. Ed.* **45,** 1744–1747 (2006).
37. Nguyen, R. V., Yao, X. & Li, C. J. Highly efficient gold-catalyzed atom-economical annulation of phenols with dienes. *Org. Lett.* **8,** 2397–2399 (2006).
38. Rosenfeld, D. C., Shekhar, S., Takemiya, A., Utsunomiya, M. & Hartwig, J. F. Hydroamination and hydroalkoxylation catalyzed by triflic acid. Parallels to reactions initiated with metal triflates. *Org. Lett.* **8,** 4179–4182 (2006).
39. Li, Z. *et al.* Bronsted acid catalyzed addition of phenols, carboxylic acids, and tosylamides to simple olefins. *Org. Lett.* **8,** 4175–4178 (2006).
40. Schwerdtfeger, P., Boyd, P. D. W., Burrell, A. K., Robinson, W. T. & Taylor, M. J. Relativistic effects in gold chemistry. 3. Gold(I) complexes. *Inorg. Chem.* **29,** 3593–3607 (1990).
41. Hertwig, R. H. *et al.* A comparative computational study of cationic coinage metal-ethylene complexes $(C_2H_4)M^+$ (M = Cu, Ag, and Au). *J. Phys. Chem.* **100,** 12253–12260 (1996).
42. Nechaev, M. S., Rayon, V. M. & Frenking, G. Energy partitioning analysis of the bonding in ethylene and acetylene complexes of Group 6, 8, and 11 metals: $(CO)_{(5)}TM\text{-}C_2H_x$ and $Cl_4TM\text{-}C_2H_x$ (TM = Cr, Mo, W), $(CO)_{(4)}TM\text{-}C_2H_x$ (TM = Fe, Ru, Os), and $TM^+\text{-}C_2H_x$ (TM = Cu, Ag, Au). *J. Phys. Chem. A* **108,** 3134–3142 (2004).
43. Fleming, I. *Frontier Orbitals and Organic Chemical Reactions* (Wiley, Chichester, 1976).
44. Cinellu, M. A. *et al.* Reactions of gold(III) oxo complexes with cyclic alkenes. *Angew. Chem. Int. Ed.* **44,** 6892–6895 (2005).
45. Hashmi, A. S. K., Weyrauch, J. P., Frey, W. & Bats, J. W. Gold catalysis: Mild conditions for the synthesis of oxazoles from *N*-propargylcarboxamides and mechanistic aspects. *Org. Lett.* **6,** 4391–4394 (2004).
46. Nevado, C. & Echavarren, A. M. Transition metal-catalyzed hydroarylation of alkynes. *Synthesis* 167–182 (2005).
47. Sromek, A. W., Rubina, M. & Gevorgyan, V. 1,2-Halogen migration in haloallenyl ketones: Regiodivergent synthesis of halofurans. *J. Am. Chem. Soc.* **127,** 10500–10501 (2005).
48. Straub, B. F. Gold(I) or gold(III) as active species in $AuCl_3$-catalyzed cyclization/cycloaddition reactions? A DFT study. *Chem. Commun.* 1726–1728 (2004).
49. Markham, J. P., Staben, S. T. & Toste, F. D. Gold(I)-catalyzed ring expansion of cyclopropanols and cyclobutanols. *J. Am. Chem. Soc.* **127,** 9708–9709 (2005).
50. Zhang, Z. *et al.* Highly active Au(I) catalyst for the intramolecular exo-hydrofunctionalization of allenes with carbon, nitrogen, and oxygen nucleophiles. *J. Am. Chem. Soc.* **128,** 9066–9073 (2006).
51. Marion, N., Díez-González, S., de Frémont, P., Noble, A. R. & Nolan, S. P. AuI-catalyzed tandem [3,3] rearrangement–intramolecular hydroarylation: Mild and efficient formation of substituted indenes. *Angew. Chem. Int. Edn* **45,** 3647–3650 (2006).
52. Sherry, B. D. & Toste, F. D. Gold(I)-catalyzed propargyl Claisen rearrangement. *J. Am. Chem. Soc.* **126,** 15978–15979 (2004).
53. Hashmi, A. S. K., Weyrauch, J. P., Rudolph, M. & Kurpejovic, E. Gold catalysis: the benefits of N and N,O ligands. *Angew. Chem. Int. Ed.* **43,** 6545–6547 (2004).
54. Zhou, C. Y., Chan, P. W. H. & Che, C. M. Gold(III) porphyrin-catalyzed cycloisomerization of allenones. *Org. Lett.* **8,** 325–328 (2006).
55. Irikura, K. K. & Goddard, W. A. Energetics of third-row transition metal methylidene ions MCH_2^+ (M = La, Hf, Ta, W, Re, Os, Ir, Pt, Au). *J. Am. Chem. Soc.* **116,** 8733–8740 (1994).
56. Heinemann, C., Hertwig, R. H., Wesendrup, R., Koch, W. & Schwarz, H. Relativistic effects on bonding in cationic transition-metal–carbene complexes—a density-functional study. *J. Am. Chem. Soc.* **117,** 495–500 (1995).
57. Barysz, M. & Pyykkö, P. Strong chemical bonds to gold. High level correlated relativistic results for diatomic $AuBe^+$, AuC^+, $AgMg^+$, and $AuSi^+$. *Chem. Phys. Lett.* **285,** 398–403 (1998).
58. Aguirre, F., Husband, J., Thompson, C. J. & Metz, R. B. Gas-phase photodissociation of $AuCH_2^+$: the dissociation threshold of jet-cooled and rotationally thermalized ions. *Chem. Phys. Lett.* **318,** 466–470 (2000).
59. Xu, Q., Imamura, Y., Fujiwara, M. & Souma, Y. A new gold catalyst: Formation of gold(I) carbonyl, $[Au(CO)_n]^+$ ($n = 1, 2$), in sulfuric acid and its application to carbonylation of olefins. *J. Org. Chem.* **62,** 1594–1598 (1997).
60. Chatt, J. & Duncanson, L. A. Olefin co-ordination compounds. 3. Infra-red spectra and structure—attempted preparation of acetylene complexes. *J. Chem. Soc.* 2939–2947 (1953).
61. Dewar, J. S. A review of the pi-complex theory. *Bull. Soc. Chim. Fr.* **18,** C71–C79 (1951).
62. deFremont, P., Scott, N. M., Stevens, E. D. & Nolan, S. P. Synthesis and structural characterization of N-heterocyclic carbene gold(I) complexes. *Organometallics* **24,** 2411–2418 (2005).
63. Raubenheimer, H. G., Esterhuysen, M. W., Timoshkin, A., Chen, Y. & Frenking, G. Electrophilic addition of Ph_3PAu^+ to anionic alkoxy Fischer-type carbene complexes: A novel approach to metal-stabilized bimetallic vinyl ether complexes. *Organometallics* **21,** 3173–3181 (2002).
64. Nakamura, I., Sato, T. & Yamamoto, Y. Gold-catalyzed intramolecular carbothiolation of alkynes: Synthesis of 2,3-disubstituted benzothiophenes from (alpha-alkoxy alkyl) (*ortho*-alkynyl phenyl) sulfides. *Angew. Chem. Int. Ed.* **45,** 4473–4475 (2006).
65. Dube, P. & Toste, F. D. Synthesis of indenyl ethers by gold(I)-catalyzed intramolecular carboalkoxylation of alkynes. *J. Am. Chem. Soc.* **128,** 12062–12063 (2006).
66. Nieto-Oberhuber, C. *et al.* Gold(I)-catalyzed cyclizations of 1,6-enynes: Alkoxycyclizations and exo/endo skeletal rearrangements. *Chem. Eur. J.* **12,** 1677–1693 (2006).
67. Nieto-Oberhuber, C. *et al.* Cationic gold(I) complexes: Highly alkynophilic catalysts for the exo- and endo-cyclization of enynes. *Angew. Chem. Int. Ed.* **43,** 2402–2406 (2004).
68. Mamane, V., Gress, T., Krause, H. & Fürstner, A. Platinum- and gold-catalyzed cycloisomerization reactions of hydroxylated enynes. *J. Am. Chem. Soc.* **126,** 8654–8655 (2004).
69. Luzung, M. R., Markham, J. P. & Toste, F. D. Catalaytic isomerization of 1,5-enynes to bicyclo[3.1.0]hexenes. *J. Am. Chem. Soc.* **126,** 10858–10859 (2004).
70. Zhang, L. & Kozmin, S. A. Gold-catalyzed assembly of heterobicyclic systems. *J. Am. Chem. Soc.* **127,** 6962–6963 (2005).
71. Fürstner, A., Stelzer, F. & Szillat, H. Platinum-catalyzed cycloisomerization reactions of enynes. *J. Am. Chem. Soc.* **123,** 11863–11869 (2001).
72. Mendez, M., Mamane, V. & Fürstner, A. Platinum-catalyzed skeletal rearrangement reactions: Generating structural diversity by a uniform mechanism. *ChemTracts Org. Chem.* **16,** 397–425 (2003).
73. Fürstner, A. & Mamane, V. Flexible synthesis of phenanthrenes by a $PtCl_2$-catalyzed cycloisomerization reaction. *J. Org. Chem.* **67,** 6264–6267 (2002).
74. Chatani, N., Inoue, H., Kotsuma, T. & Murai, S. Skeletal reorganization of enynes to 1-vinylcycloalkenes catalyzed by $GaCl_3$. *J. Am. Chem. Soc.* **124,** 10294–10295 (2002).
75. Mamane, V., Hannen, P. & Fürstner, A. Synthesis of phenanthrenes and polycyclic heteroarenes by transition-metal catalyzed cycloisomerization reactions. *Chem. Eur. J.* **10,** 4556–4575 (2004).
76. Christian, B. Electrophilic activation and cycloisomerization of enynes: A new route to functional cyclopropanes. *Angew. Chem. Int. Ed.* **44,** 2328–2334 (2005).
77. Ma, S., Yu, S. & Gu, Z. Gold-catalyzed cyclization of enynes. *Angew. Chem. Int. Ed.* **45,** 200–203 (2006).

78. Furstner, A., Davies, P. W. & Gress, T. Cyclobutenes by platinum-catalyzed cycloisomerization reactions of enynes. *J. Am. Chem. Soc.* **127,** 8244–8245 (2005).
79. Oi, S., Tsukamoto, I., Miyano, S. & Inoue, Y. Cationic platinum-complex-catalyzed skeletal reorganization of enynes. *Organometallics* **20,** 3704–3709 (2001).
80. BhanuPrasad, B. A., Yoshimoto, F. K. & Sarpong, R. Pt-catalyzed pentannulations from in situ generated metallo-carbenoids utilizing propargylic esters. *J. Am. Chem. Soc.* **127,** 12468–12469 (2005).
81. Ito, Y., Sawamura, M. & Hayashi, T. Catalytic asymmetric aldol reaction—reaction of aldehydes with isocyanoacetate catalyzed by a chiral ferrocenylphosphine-gold(I) complex. *J. Am. Chem. Soc.* **108,** 6405–6406 (1986).
82. Munoz, M. P., Adrio, J., Carretero, J. C. & Echavarren, A. M. Ligand effects in gold- and platinum-catalyzed cyclization of enynes: Chiral gold complexes for enantioselective alkoxycyclization. *Organometallics* **24,** 1293–1300 (2005).
83. Shi, X., Gorin, D. J. & Toste, F. D. Synthesis of 2-cyclopentenones by gold(I)-catalyzed Rautenstrauch rearrangement. *J. Am. Chem. Soc.* **127,** 5802–5803 (2005).
84. Rautenstrauch, V. 2-Cyclopentenones from 1-ethynyl-2-propenyl acetates. *J. Org. Chem.* **49,** 950–952 (1984).
85. Faza, O. N., Lopez, C. S., Alvarez, R. & de Lera, A. R. Mechanism of the gold(I)-catalyzed Rautenstrauch rearrangement: A center-to-helix-to-center chirality transfer. *J. Am. Chem. Soc.* **128,** 2434–2437 (2006).
86. Fehr, C. & Galindo, J. Synthesis of (−)-cubebol by face-selective platinum-, gold-, or copper-catalyzed cycloisomerization: Evidence for chirality transfer. *Angew. Chem. Int. Edn* **45,** 2901–2904 (2006).
87. Furstner, A. & Hannen, P. Carene terpenoids by gold-catalyzed cycloisomerization reactions. *Chem. Commun.* 2546–2547 (2004).
88. Furstner, A. & Hannen, P. Platinum- and gold-catalyzed rearrangement reactions of propargyl acetates: Total syntheses of (−)-alpha-cubebene, (−)-cubebol, sesquicarene and related terpenes. *Chem. Eur. J.* **12,** 3006–3019 (2006).
89. Johansson, M. J., Gorin, D. J., Staben, S. T. & Toste, F. D. Gold(I)-catalyzed stereoselective olefin cyclopropanation. *J. Am. Chem. Soc.* **127,** 18002–18003 (2005).
90. Miki, K., Ohe, K. & Uemura, S. A new ruthenium-catalyzed cyclopropanation of alkenes using propargylic acetates as a precursor of vinylcarbenoids. *Tetrahedr. Lett.* **44,** 2019–2022 (2003).
91. Gorin, D. J., Davis, N. R. & Toste, F. D. Gold(I)-catalyzed intramolecular acetylenic Schmidt reaction. *J. Am. Chem. Soc.* **127,** 11260–11261 (2005).
92. Hashmi, A. S. K., Blanco, M. C., Kurpejovic, E., Frey, W. & Bats, J. W. Gold catalysis: First applications of cationic binuclear gold(I) complexes and the first intermolecular reaction of an alkyne with a furan. *Adv. Synth. Catal.* **348,** 709–713 (2006).
93. Fructos, M. R. *et al.* A gold catalyst for carbene-transfer reactions from ethyl diazoacetate. *Angew. Chem. Int. Ed.* **44,** 5284–5288 (2005).
94. Hoffmann-Roder, A. & Krause, N. The golden gate to catalysis. *Org. Biomol. Chem.* **3,** 387–391 (2005).
95. Hashmi, A. S. K. The catalysis gold rush: New claims. *Angew. Chem. Int. Ed.* **44,** 6990–6993 (2005).
96. Pitzer, K. S. Relativistic effects on chemical properties. *Acc. Chem. Res.* **12,** 272–276 (1979).
97. Pyykkö, P. Relativistic effects in structural chemistry. *Chem. Rev.* **88,** 563–594 (1988).
98. Norrby, L. J. Why is mercury liquid—or, why do relativistic effects not get into chemistry textbooks? *J. Chem. Educ.* **68,** 110–113 (1991).
99. Bagus, P. S., Lee, Y. S. & Pitzer, K. S. Effects of relativity and of lanthanide contraction on atoms from hafnium to bismuth. *Chem. Phys. Lett.* **33,** 408–411 (1975).
100. Desclaux, J. P. Relativistic Dirac–Fock expectation values for atoms with $Z = 1$ to $Z = 120$. *Atom. Data Nucl. Data Tables* **12,** 311–406 (1973).

Acknowledgements We thank P. Pyykkö for discussions during the preparation of this manuscript. Funding from the University of California, Berkeley, NIHGMS, Merck Research Laboratories, Bristol-Myers Squibb, Amgen Inc., DuPont, GlaxoSmithKline, Eli Lilly & Co., Pfizer, AstraZeneca, Abbott, Boehringer Ingelheim, Novartis and Roche is gratefully acknowledged. D.J.G. is an ACS Organic Division predoctoral fellow.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to F.D.T. (fdtoste@berkeley.edu).

Vol 446|22 March 2007|doi:10.1038/nature05569

ARTICLES

Total synthesis of marine natural products without using protecting groups

Phil S. Baran[1], Thomas J. Maimone[1] & Jeremy M. Richter[1]

The field of organic synthesis has made phenomenal advances in the past fifty years, yet chemists still struggle to design synthetic routes that will enable them to obtain sufficient quantities of complex molecules for biological and medical studies. Total synthesis is therefore increasingly focused on preparing natural products in the most efficient manner possible. Here we describe the preparative-scale, enantioselective, total syntheses of members of the hapalindole, fischerindole, welwitindolinone and ambiguine families, each constructed without the need for protecting groups—the use of such groups adds considerably to the cost and complexity of syntheses. As a consequence, molecules that have previously required twenty or more steps to synthesize racemically in milligram amounts can now be obtained as single enantiomers in significant quantities in ten steps or less. Through the extension of the general principles demonstrated here, it should be possible to access other complex molecular architectures without using protecting groups.

Although the field of total synthesis[1,2] has made great advances since 1828 (ref. 3), it is still far from being a mature or applied science[4,5]. For example, precise control over the individual reactivity of functional groups within a complex molecular architecture (chemoselectivity) still remains a largely unanswered challenge. Historically, the use of protecting groups has been the standard solution to this problem because they allow functional groups to be dealt with on an individual basis. Indeed, these functionality masks have permeated organic chemistry to the extent that textbooks state that avoiding them is impossible[6,7]. Their use has become routine even on molecules of low complexity[8]. Ideally, protecting groups are easily appended, allow one to smoothly perform the initially intended transformation, and then gracefully depart without incident. In practice, however, these artificial devices add at least two steps each to a synthetic sequence and sometimes dramatically lower the efficiency of a synthesis owing to unforeseen difficulties encountered during their removal or unintended side reactions initiated by their presence[9]. Ironically, their presence can lead to an additional layer of chemoselectivity considerations that often take centre stage within a complex total synthesis endeavour[8]. The multitude of complications imparted by protecting-group manipulations contributes to the perception that natural products, despite their overwhelming utility in medicine, are too complex to be synthesized efficiently in a drug discovery setting[10,11].

Figure 1 summarizes three different approaches to chemical synthesis using the complex natural product ambiguine H (**1**) as an example. In a biological setting, where the goal of synthesis is to create function rather than a specific target molecule, simple feedstock chemicals are woven together without protecting groups by using exquisitely selective enzymes[12]. For instance, it has been proposed that the key C–C bonds of the ambiguines are forged from an enzymatic enantioselective cation-olefin cyclization of a simple hydrocarbon with a 3-substituted indole[13]. Indeed, emulating nature (biomimetic synthesis) can sometimes lead to extremely efficient synthetic routes[1,2,14–16]. In contrast, a standard approach to synthesis uses strategic disconnections that are often made in order to shield perceived functional group incompatibilities *en route* to a specific target. Here we describe syntheses, the inspiration for which comes partly from studying the biosynthetic pathway, strictly avoid the use of protecting groups, and harness the natural reactivity of specific functional groups within a complex setting. This approach has led to solutions that would not have been apparent had the natural tendencies of the reactive centres been masked.

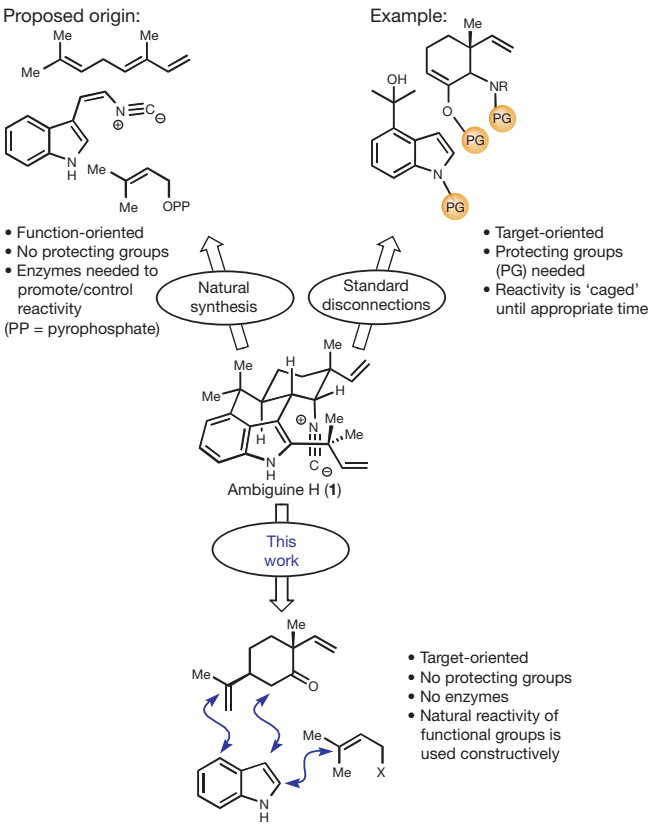


Figure 1 | Approaches to chemical synthesis. Here we show ambiguine H (**1**) as an example.

[1]Department of Chemistry, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California 92037, USA.

The Stigonemataceae family of cyanobacteria has produced a class of over 60 biogenetically related, architecturally complex, topologically unique, and functionally rich indole natural products that form the basis of the hapalindole[17,18], fischerindole[13], welwitindolinone[13], and ambiguine[19,20] alkaloids (Fig. 2). They exhibit a broad range of bioactivities including antifungal, antibacterial, antimycotic and anticancer properties, with some members having potencies comparable to clinical agents (streptomycin, puramycin and amphotericin)[13,17–20]. Further study of these potential medicinal agents is hampered by the fact that the cyanobacteria produce complex mixtures of these natural products in low yield. For instance, small quantities (about 5 mg) of **1**, **2**, **4** and **5** have been isolated in yields ranging from 0.00671% (for **2**) to 0.0213% (for **5**) following tedious purification and HPLC separation.

Total synthesis of hapalindole U (2) and ambiguine H (1)

Although there have been no published synthetic routes to the ambiguines, racemic hapalindole U (**2**, Fig. 2) has been constructed in 20 steps by a non-stereocontrolled sequence with multiple protecting groups[21]. Fig. 3 outlines a simple, enantioselective entry to the ambiguine alkaloid family, by way of **2**, facilitated by newly developed methodology for C–C bond formation and a deliberate effort to eliminate the use of protecting groups.

The synthesis commenced with readily available terpene **7**, which is synthesized in four steps by a route that closely parallels the strategy of ref. 22 (see Supplementary Information). The indole and terpene subunits were then merged without protecting groups using a direct indole coupling, a reaction that was invented specifically for forming this type of C–C bond[23]. This reaction has already been successfully employed in short syntheses of **3** and **6** (Fig. 2)[23] and furnished indole **8** as a single diastereomer in 61% isolated yield. An extensive screening of acids for the requisite site-selective (carbon C4) Friedel-Crafts annulation failed to furnish detectable quantities of **10**. Not surprisingly, cyclization at C2 rather than C4 of the indole was observed, along with decomposition. Rather than resorting to protecting groups to either shield the C2 position or as a means of tuning the electronic nature of the indole, a different strategy was pursued using an indole building block biased to react at the C4 position. The brominated indole **9** was therefore targeted as a potential precursor to **10** via a radical- or transition-metal-mediated cyclization. As a testament to the versatility of the direct indole coupling, 4-bromoindole merged with **7** reliably on a gram scale to produce **9** in 50% isolated yield. We note that this mode of C–C bond formation is orthogonal to other transition-metal-mediated processes in that coupling occurs selectively by C–H bond functionalization rather than by C–Br bond insertion. This reaction is reliant on the presence of a free N–H[24] and so we would not have discovered it had we resorted to protecting groups.

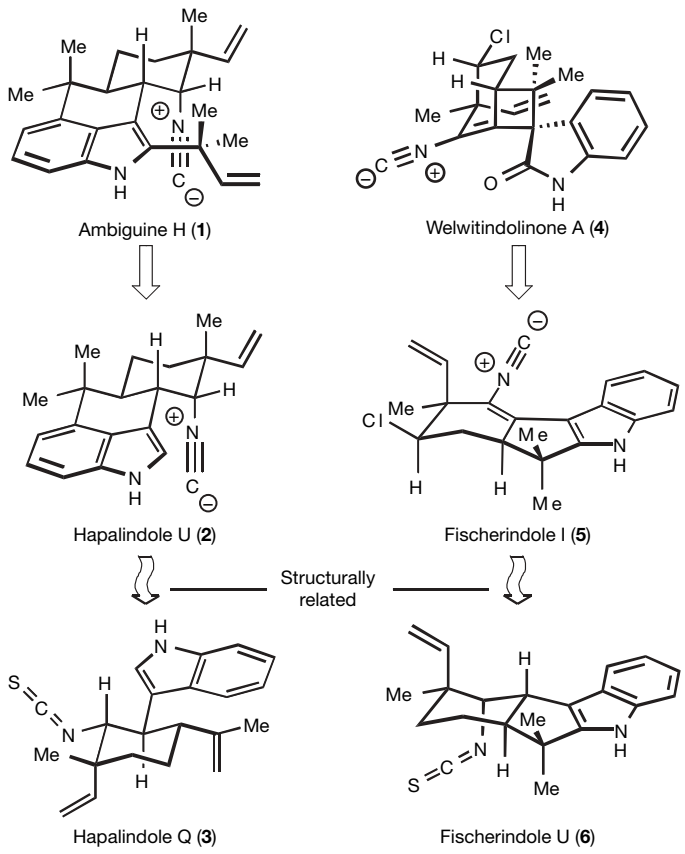


Figure 2 | Representative members of the ambiguine, fischerindole, hapalindole and welwitindolinone alkaloid families and proposed biosynthetic relationships.

To elicit the desired 6-*exo*-trig cyclization of **9** to **10**, radical- and palladium-mediated methods were explored. Although the former led mainly to the undesired 7-*endo*-trig and debrominated products, the reductive Heck[25] methods of Larock[26] and Grigg[27] showed some promise (18–39% isolated yield of **10** with significant amounts of **8**). Many conditions were screened to maximize the conversion of **9** to the annulation product **10** while suppressing the competing debromination pathway, leading to the formation of **8** as well as destruction of the catalyst in the highly reducing environment (see Supplementary Information for details). We discovered that the use of Herrmann's catalyst[28] as the Pd-source and its slow addition were necessary to reliably achieve a 65% isolated yield of **10**.

With the tetracyclic core of the ambiguines and hapalindoles in place, ketone **10** could be easily converted to hapalindole U (**2**) by stereocontrolled, microwave-assisted reductive amination, followed by formylation of the crude amine and dehydration of the resulting formamide. The overall isolated yield for the two-pot sequence was 60%. Synthetic **2**, prepared in four steps from ketone **7** (20% isolated yield, >1 g prepared), was spectroscopically identical to that reported by Moore[18] and was confirmed by X-ray crystallography: melting point 241 °C (decomposition, dec.), hexanes:Et_2O:MeOH, 10:5:1.

All that remained to bridge the gap between the hapalindole and ambiguine families (**2** → **1**) was the seemingly straightforward task of installing the *tert*-prenyl unit onto C2 of **2**. However, our attempts to achieve direct or indirect *tert*-prenylation onto **2**, as well as earlier intermediates, were unsuccessful owing to the incompatibility of the isonitrile with acids and transition metals, as well as the unusual reactivity of the indole nucleus within the tetracyclic ring system. For example, attempts to activate the indole for nucleophilic addition at carbon C2 (that is, C3 chloroindolenine formation) always led to either functionalization of C2 with the activating agent (via [1,2]-shift), attack at nitrogen N1 (with expulsion of the activating agent), or attack at the activating agent (returning starting material). On the basis of these empirical observations, and instead of resorting to protecting group chemistry to shield the fragile isonitrile and indole N–H, we devised a strategy to accommodate and exploit the natural reactivity of both functionalities.

Thus, exposure of **2** to *tert*-BuOCl, followed by prenyl 9-BBN, according to Danishefsky's protocol[29], produced the unusual crystalline pentacyclic chloroimidate **12** (structure confirmed by X-ray spectroscopy: melting point 244 °C (dec.), Et_2O). This product is presumably formed by a tandem sequence involving initial chlorination of the axial-configured isonitrile, nucleophilic attack of indole (C3), and addition of the prenyl reagent to the resulting imine **11**. The observed stereochemistry of the *tert*-prenyl unit at C2 probably stems from its addition to the less-hindered face of the folded architecture of imine **11**. The unorthodox nature of this transformation is consistent with the individual reactivity observed previously for the isonitrile and indole (see above). We reasoned that a Norrish-type cleavage[30] of the chloroimidate in **12** might initiate a fragmentation cascade to liberate the BBN functionality, the

extraneous chlorine atom, an unwanted C–C bond, and restore the indole and isocyanide moieties. Indeed, irradiation of **12** for five hours led to ambiguine H (**1**), accomplishing all five necessary tasks in a single step (63% yield based on recovered **12**). We suggest a mechanism for this transformation in Fig. 3. If the reactive functionalities of **2** were shielded with protecting groups, such chemical reactivity would not have been apparent (that is, the Norrish-like cleavage of a chloroimidate or the use of a sensitive isonitrile to assist in the activation of a free indole). Synthetic **1** exhibited identical spectroscopic data to that reported in ref. 20 and was confirmed by X-ray crystallography (melting point 228–231 °C (dec.), hexanes/Et_2O (1:1)), representing the first total synthesis of a member of the ambiguine natural product family. Because **1** is unstable on prolonged storage, we made gram quantities of **2** and converted it to **1** as needed; see Supplementary Information for details.

Total synthesis of welwitindolinone A (4) and fischerindole I (5)

The elimination of protecting groups and reduction of the number of steps in a total synthesis can also simplify the optimization of the overall yield of a sequence. Statistics dictate that because each step in a shorter sequence carries a greater impact on the overall efficiency of a synthesis, optimization is realized more rapidly than with the corresponding longer routes[31]. The recent total syntheses of fischerindole I (**4**) and welwitindolinone A (**5**)[32] are an illustration of this point. Although they represent some of the most complex natural products to be synthesized without protecting groups[8] and required only seven to eight chemical operations, their syntheses had overall yields of only 6.9% and 1.7%, respectively, from ketone **15** (Fig. 4). The synthesis also suffered from limited scalability owing to the technically demanding nature of the final two steps of the sequence. Figure 4 depicts revised syntheses of **4** and **5** that can be conducted on a much larger scale than that reported previously and in overall yields of 13.0% and 5.7%, respectively, from **15**—via optimization of individual steps, not an alteration in general strategy.

In five simple steps from carvone oxide[32], amine **17** is accessible in large quantities via the direct coupling of chloroketone **15** with indole (62% yield) followed by Friedel–Crafts cyclization and stereocontrolled reductive amination of **16** (see Supplementary Information for details). Amine **17** is then formylated, followed by immediate dehydration with phosgene to install the isonitrile functionality and furnish 11-*epi*-fischerindole G (**18**). In our previous route, **18** provided a scaffold on which to perform the requisite unsaturation to form **5** and an ensuing ring contraction to form **4**. We reasoned that the yield and selectivity problems in that route stemmed from the choice of a chlorine-based oxidant (*tert*-BuOCl) that was both inefficient and unselective. As shown in the synthesis of ambiguine H (see below), such oxidants react readily with isonitriles. To accomplish the conversion of **18** to **5**, an oxidant was chosen that was more suited to benzylic oxidations. By simply exposing **18** to DDQ[33] in the presence of water, fischerindole I (**5**) was produced in excellent overall yield (>2 g prepared), presumably through the intermediate α,β-unsaturated imine **19**. For the ensuing oxidative ring contraction, we reasoned that a hitherto-unknown fluorohydroxylation of indole rather than chlorohydroxylation should suppress isonitrile-derived side-product formation, owing to the increased hardness of fluorine over chlorine. A method for fluorohydroxylation of the indole moiety in **5** was developed using xenon difluoride[34] and water in acetonitrile, providing welwitindolinone A scaleably (>390 mg prepared), in 44% isolated yield, and as a single diastereomer. This cascade sequence can be envisioned to proceed through fluorination of the indole nucleus to give **20**, which is trapped with water to give **21**. Elimination of fluoride would give the azaorthoquinodimethane (**22**)[35], which undergoes a [1,5] sigmatropic rearrangement to furnish the spirocyclobutane of welwitindolinone A, as a single diastereomer. The observed chemoselectivity (in the presence of two other olefins and a reactive isonitrile) of this new reaction is worthy of further study. Because **4** and **5** are unstable on prolonged storage, gram quantities of **18** are made and converted to **4** and **5** as needed (see Supplementary

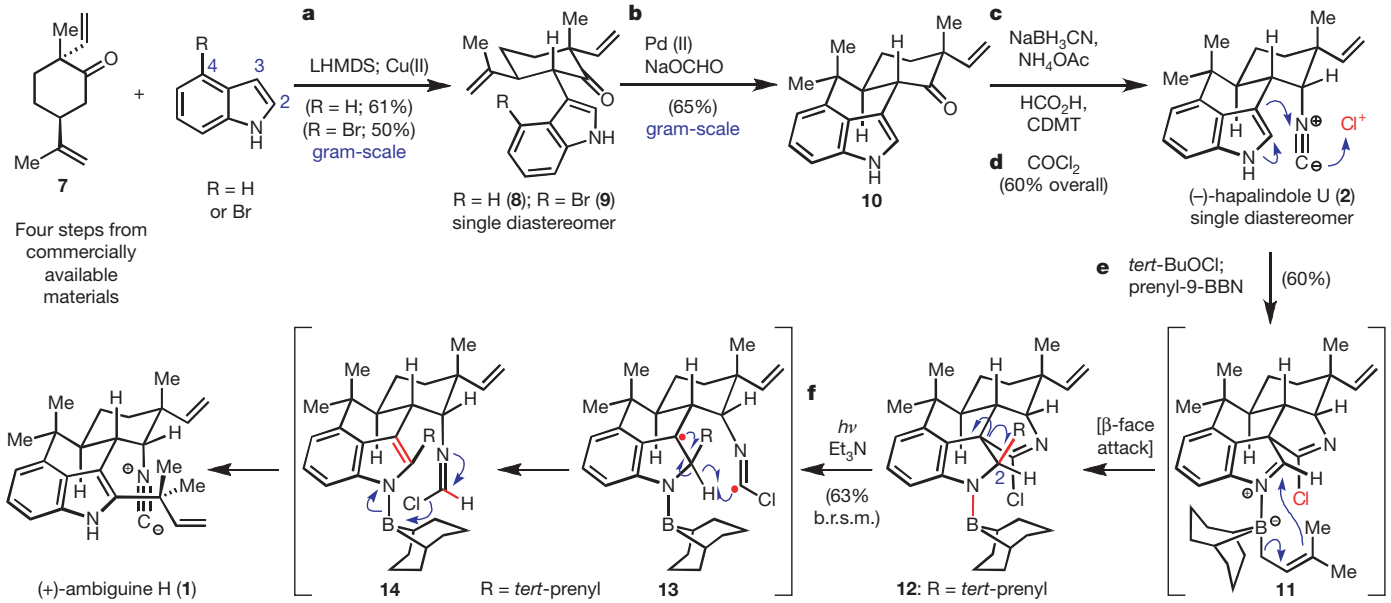


Figure 3 | Protecting-group-free synthesis of ambiguine H (1) and hapalindole U (2). Reagents and conditions as follows. **a**, Indole (1.9 equiv.), ketone **7** (1.0 equiv.), LHMDS (3.4 equiv.), Cu(II)-2-ethylhexanoate (1.5 equiv.), THF, starting temperature −78 °C, 5 min to 25 °C, yield is 61%; or 4-bromoindole (2.8 equiv.), ketone **7** (1.0 equiv.), LHMDS (4.4 equiv.), Cu(II)-2-ethylhexanoate (2.0 equiv.), THF, –78 °C, 5 min to 25 °C, 50%. **b**, $[Pd(P(o\text{-}tol)_3OAc]_2$ (0.05 equiv.), NaOCHO (1.25 equiv.), TBAB (2.0 equiv.), Et_3N (2.2 equiv.), DMF, 80 °C, slow addition of Pd over 5 h, 65%. **c**, NH_4OAc (40 equiv.), $NaCNBH_3$ (9.3 equiv.), MeOH/THF, microwave irradiation at 150 °C, 2.5 min; then HCO_2H (2.0 equiv.), CDMT (2.2 equiv.), DMAP (0.05 equiv.), NMM (2.2 equiv.), DCM, 2 h, 25 °C. **d**, $COCl_2$ (2.0 equiv.), Et_3N (17.5 equiv.), DCM, 0 °C, 60% over two steps. **e**, *tert*-BuOCl (1.15 equiv), DCM, –78 °C, 12 min; then prenyl-9-BBN (2.0 equiv), −78 °C, 30 min, 60%. **f**, Et_3N (5.0 equiv.), benzene, *hν*, 5 h, 63% b.r.s.m. (based on recovered starting material). LHMDS, lithium hexamethyldisilazide; THF, tetrahydrofuran; TBAB, *tetra-n*-butyl ammonium bromide; Et_3N, triethylamine; DMF, *N*,*N*-dimethylformamide; CDMT, 2-chloro-4,6-dimethoxy-1,3,5-triazine; DMAP, 4-*N*-dimethylaminopyridine; NMM, *N*-methylmorpholine; DCM, dichloromethane; 9-BBN, 9-borabicyclo-nonane. For selected physical data for compounds **1**, **2**, **7**–**10** and **12**, see the Supplementary Information. Compounds **2**, **12** and **1** were verified by X-ray crystallography.

Information for details). The revised routes to **4** and **5** demonstrate how such mechanistically inspired reagent changes can greatly improve the overall efficiency of an extremely short synthesis. The only other reported total synthesis of **4** requires 25 steps and six protecting groups to deliver milligram quantities of racemic material[36].

Discussion

Taken together with the concepts of ''atom economy''[37] and ''step economy,''[38] we followed several general guidelines during the planning stage (retrosynthetic analysis)[39] of these syntheses: (1) redox reactions that do not form C–C bonds should be minimized[40], (2) the percentage of C–C bond-forming events within the total number of steps in a synthesis should be maximized[39,40], (3) disconnections should be made to maximize convergency[41], (4) the overall oxidation level of intermediates should linearly escalate during assembly of the molecular framework (except in cases where there is strategic benefit such as an asymmetric reduction)[42], (5) where possible, cascade (tandem) reactions should be designed and incorporated to elicit maximum structural change per step[43], (6) the innate reactivity of functional groups should be exploited so as to reduce the number of (or perhaps even eliminate) protecting groups[8,44], (7) effort should be spent on the invention of new methodology to facilitate the aforementioned criteria and to uncover new aspects of chemical reactivity[45], (8) if the target molecule is of natural origin, biomimetic pathways (either known or proposed) should be incorporated to the extent that they aid the above considerations[1,2,14–16,46]. Although these principles have existed conceptually and separately for several years[1,2,8,47], this series of total syntheses cohesively applies them as a whole.

Despite the demonstrated advantages, there are some limitations to deliberately excluding protecting groups from the synthesis of complex molecules. For instance, their inclusion within a synthetic plan may allow for a certain level of security, because perceived functional-group incompatibilities can be dealt with at the outset. Indeed, omitting protecting groups during the retrosynthetic planning stages of a complex molecule might involve a certain amount of risk and speculation, owing to the unpredictable reactivity that is inevitably encountered at the late stages of a total synthesis[48]. In some cases, the use of protecting groups may offer a more efficient or even the sole solution. For example, the total synthesis of certain classes of molecules, such as poly-ketides, -peptides, -saccharides, and -nucleotides, will perhaps always require some level of protection (not only owing to a lack of chemoselectivity but also the practical issues of purification and characterization).

In summary, representative members of a large class of natural products consisting of four different families have been constructed by adhering to the general principles outlined above. The enantioselective total syntheses of ambiguine H (**1**), hapalindole U (**2**), welwitindolinone A (**4**), and fischerindole I (**5**) require only seven to ten steps from commercially available materials and can easily be performed on a preparative scale using inexpensive reagents. Of those steps, approximately half involved C–C bond formation and aside from a stereoselective reductive amination, the oxidation states of intermediates gradually escalated from beginning to end. Certain aspects of these convergent syntheses also benefited from insights into their biosynthetic origins and the incorporation of designed cascade reactions. Finally, the deliberate exclusion of protecting groups from the overall synthetic design facilitated the development and discovery of new chemical reactions by harvesting the intrinsic reactivity within organic molecules.

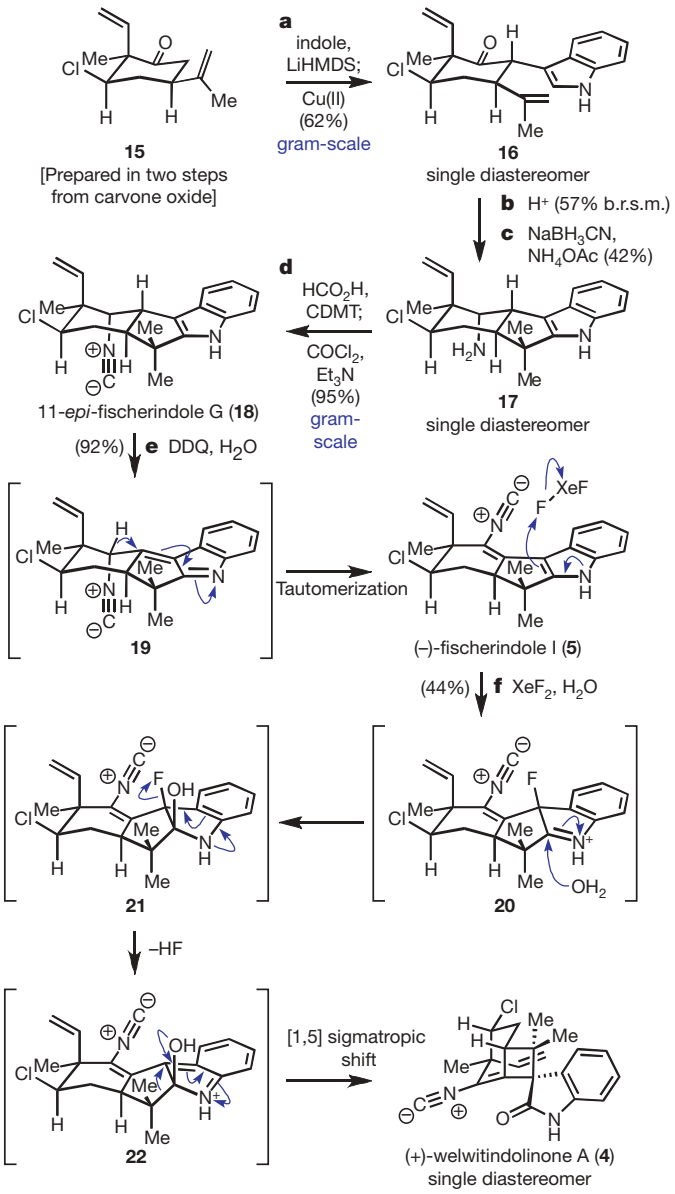


Figure 4 | Protecting-group-free total synthesis of fischerindole I (5) and welwitindolinone A (4). Reagents and conditions as follows. **a**, Indole (2.0 equiv.), LiHMDS (3.3 equiv.), THF, −78 °C, 30 min, copper(II)-2-ethylhexanoate (1.5 equiv.), −78 to 23 °C, 20 min, 62%. **b**, Montmorillonite K-10 clay, microwave irradiation at 120 °C, 6 min, 57% b.r.s.m. **c**, NH_4OAc (40 equiv.), $NaCNBH_3$ (7.5 equiv.), 3 Å molecular sieves, MeOH/THF, sonication, 18 h, 42%. **d**, HCO_2H (2.0 equiv.), CDMT (2.2 equiv.), DMAP (0.1 equiv.), NMM (2.2 equiv.), DCM, 23 °C, 30 min; Et_3N (17.5 equiv.), $COCl_2$ (2.0 equiv.), DCM, 0 °C, 10 min, 95%. **e**, DDQ (2.5 equiv.), H_2O, THF, 0 °C, 30 min, 92%. **f**, XeF_2, H_2O, MeCN, 23 °C, 5 min; 44%. DDQ, 2,3-dichloro-5,6-dicyanobenzoquinone; MeCN, acetonitrile. For selected physical data for compounds **18**, **5** and **4** see the Supplementary Information.

METHODS

All reactions were carried out under a nitrogen atmosphere with dry solvents under anhydrous conditions, unless otherwise noted. Yields refer to chromatographically and spectroscopically homogenous materials, unless otherwise stated. Reagents were purchased at the highest commercial quality and used without further purification, unless otherwise stated. Reactions were monitored by thin-layer chromatography. For full experimental details and procedures for all reactions performed and full characterization (^1H and ^{13}C nuclear magnetic resonance, high-resolution mass spectrometry, infrared, optical rotation, melting point, and R_f value) of all new compounds, see the Supplementary Information.

Received 11 October 2006; accepted 2 January 2007.

1. Nicolaou, K. C. & Sorensen, E. J. *Classics in Total Synthesis* (VCH, New York, 1996).
2. Nicolaou, K. C. & Snyder, S. A. *Classics in Total Synthesis II* (Wiley-VCH, Weinheim, 2003).
3. Wöhler, F. Ueber die Künstliche Bildung des Harnstoffe Poggendorfs. *Ann. Phys. Chem.* **12**, 253–256 (1828).
4. Service, R. F. Race for molecular summits. *Science* **285**, 184–187 (1999).

5. Seebach, D. Organic synthesis–where now? *Angew. Chem. Int. Edn Engl.* **29,** 1320–1367 (1990).
6. Kocienski, P. J. *Protecting Groups* 3rd edn (Thieme, New York, 2005).
7. Green, T. W. & Wuts, P. G. *Protective Groups in Organic Synthesis* 3rd edn (Wiley, Hoboken, 1999).
8. Hoffmann, R. W. Protecting group free synthesis. *Synthesis* 3531–3541 (2006).
9. Sierra, M. A. & de la Torre, M. C. *Dead Ends and Detours, Direct Ways to Successful Total Synthesis* (Wiley-VCH, Weinheim, 2004).
10. Butler, M. S. The role of natural product chemistry in drug discovery. *J. Nat. Prod.* **67,** 2141–2153 (2004).
11. Wilson, R. M. & Danishefsky, S. J. Small molecule natural products in the discovery of therapeutic agents: the synthesis connection. *J. Org. Chem.* **71,** 8329–8351 (2006).
12. Faber, K. *Biotransformations in Organic Chemistry* 3rd edn (Springer, New York, 1997).
13. Stratmann, K. *et al.* Welwitindolinones, unusual alkaloids from the blue-green algae *Hapalosiphon welwitschii* and *Westiella intricate*. Relationship to fischerindoles and hapalindoles. *J. Am. Chem. Soc.* **116,** 9935–9942 (1994).
14. Scholz, U. & Winterfeldt, E. Biomimetic synthesis of alkaloids. *Nat. Prod. Rep.* **17,** 349–366 (2000).
15. Eschenmoser, A. Vitamin B_{12}: experiments concerning the origin of its molecular structure. *Angew. Chem. Int. Edn Engl.* **27,** 5–39 (1988).
16. Heathcock, C. H. The enchanting alkaloids of Yuzuriha. *Angew. Chem. Int. Edn Engl.* **31,** 665–681 (1992).
17. Moore, R. E., Cheuk, C. & Patterson, G. M. L. Hapalindoles: new alkaloids from the blue-green alga *Hapalosiphon fontinalis. J. Am. Chem. Soc.* **106,** 6456–6457 (1984).
18. Moore, R. E. *et al.* Hapalindoles, antibacterial and antimycotic alkaloids from the cyanophyte *Hapalosiphon fontinalis. J. Org. Chem.* **52,** 1036–1043 (1987).
19. Smitka, T. A. *et al.* Ambiguine isonitriles, fungicidal hapalindole-type alkaloids from three genera of blue-green algae belonging to the Stigonemataceae. *J. Org. Chem.* **57,** 857–861 (1992).
20. Raveh, A. & Carmeli, S. Antimicrobial ambiguines from the cyanobacterium *Fischerella* sp. collected in Israel. *J. Nat. Prod.* doi:10.1021/np060495r (in the press); published online 13 January 2007.
21. Muratake, H., Kumagami, H. & Natsume, M. Synthetic studies of marine alkaloids hapalindoles. Part 3. Total synthesis of (±)-hapalindoles H and U. *Tetrahedron* **46,** 6351–6360 (1990).
22. Mehta, G. & Acharyulu, P. V. R. Terpenes to terpenes. Stereo- and enantio-selective synthesis of (+)-α-elemene and a short route to a versatile diquinane chiron. *J. Chem. Soc. Chem. Commun.* 2759–2760 (1994).
23. Baran, P. S. & Richter, J. M. Direct coupling of indoles with carbonyl compounds: short, enantioselective, gram-scale synthetic entry into the hapalindole and fischerindole alkaloid families. *J. Am. Chem. Soc.* **126,** 7450–7451 (2004).
24. Baran, P. S., Richter, J. M. & Lin, D. W. Direct coupling of pyrroles with carbonyl compounds: short enantioselective synthesis of (*S*)-ketorolac. *Angew. Chem. Int. Edn Engl.* **44,** 609–612 (2005).
25. Nicolaou, K. C., Bulger, P. G. & Sarlah, D. Palladium-catalyzed cross-coupling reactions in total synthesis. *Angew. Chem. Int. Edn Engl.* **44,** 4442–4489 (2005).
26. Larock, R. C. & Babu, S. Synthesis of nitrogen heterocycles via palladium-catalyzed intramolecular cyclization. *Tetrahedr. Lett.* **28,** 5291–5294 (1987).
27. Burns, B. *et al.* Palladium catalysed tandem cyclisation-anion capture processes. Part 1: Background and hydride ion capture by alkyl- and π-allyl-palladium species. *Tetrahedr. Lett.* **29,** 4329–4332 (1988).
28. Herrmann, W. A. *et al.* Palladacycles as structurally defined catalysts for the Heck olefination of chloro- and bromoarenes. *Angew. Chem. Int. Edn Engl.* **34,** 1844–1848 (1995).
29. Schkeryantz, J. M., Woo, J. C. G., Siliphaivanth, P., Depew, K. M. & Danishefsky, S. J. Total synthesis of gypsetin, deoxybrevianamide E, brevianamide E, and tryprostatin B: Novel constructions of 2,3-disubstituted indoles. *J. Am. Chem. Soc.* **121,** 11964–11975 (1999).
30. Turro, N. J. *Modern Molecular Photochemistry* Ch. 13 (University Science Books, Sausalito, 1991).
31. Hudlicky, T. Design constraints in practical syntheses of complex molecules: current status, case studies with carbohydrates and alkaloids, and future perspectives. *Chem. Rev.* **96,** 3–30 (1996).
32. Baran, P. S. & Richter, J. M. Enantioselective total synthesis of welwitindolinone A and fischerindoles I and G. *J. Am. Chem. Soc.* **127,** 15394–15396 (2005).
33. Oikawa, Y. & Yonemitsu, O. Selective oxidation of the side chain at C–3 of indoles. *J. Org. Chem.* **42,** 1213–1216 (1977).
34. Shellhamer, D. F. *et al.* Reaction of xenon difluoride with indene in aqueous 1,2-dimethoxyethane and tetrahydrofuran. *J. Chem. Soc. Perkin Trans.* **II,** 401–403 (1991).
35. Baran, P. S. & Shenvi, R. A. Total synthesis of (±)-chartelline C. *J. Am. Chem. Soc.* **128,** 14028–14029 (2006).
36. Reisman, S. E., Ready, J. M., Hasuoka, A., Smith, C. J. & Wood, J. L. Total synthesis of (±)-welwitindolinone A isonitrile. *J. Am. Chem. Soc.* **128,** 1448–1449 (2006).
37. Trost, B. M. The atom economy—a search for synthetic efficiency. *Science* **254,** 1471–1477 (1991).
38. Wender, P. A. & Miller, B. L. in *Organic Synthesis: Theory and Applications* (ed. Hudlicky, T.) Vol. 2, 27–66 (JAI Press, Greenwich, Connecticut, 1993).
39. Corey, E. J. & Cheng, X.-M. *The Logic of Chemical Synthesis* (Wiley, New York, 1995).
40. Hendrickson, J. B. Systematic synthesis design. IV. Numerical codification of construction reactions. *J. Am. Chem. Soc.* **97,** 5784–5800 (1975).
41. Bertz, S. H. Convergence, molecular complexity, and synthetic analysis. *J. Am. Chem. Soc.* **104,** 5801–5803 (1982).
42. Evans, D. A. *Synthesis Design and the Oxidation State Issue* (Lecture at Scripps Research Institute, 12 February, 2004).
43. Nicolaou, K. C., Edmonds, D. J. & Bulger, P. G. Cascade reactions in total synthesis. *Angew. Chem. Int. Edn Engl.* **45,** 7134–7186 (2006).
44. Hoveyda, A. H., Evans, D. A. & Fu, G. C. Substrate-directable chemical reactions. *Chem. Rev.* **93,** 1307–1370 (1993).
45. Wender, P. A. (ed.) Frontiers in Organic Synthesis. *Chem. Rev.* **96** (special issue), 1–600 (1996).
46. Vanderwal, C. D., Vosburg, D. A., Weiler, S. & Sorensen, E. J. An enantioselective synthesis of FR182877 provides a chemical rationalization of its structure and affords multigram quantities of its direct precursor. *J. Am. Chem. Soc.* **125,** 5393–5407 (2003).
47. Nicolaou, K. C. & Snyder, S. A. The essence of total synthesis. *Proc. Natl Acad. Sci. USA* **101,** 11929–11936 (2004).
48. Dörwald, F. Z. *Side Reactions in Organic Synthesis* (Wiley-VCH, Weinheim, 2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to M. R. Ghadiri for discussions and comments on the manuscript, P. Mariano for his comments on the mechanism of photocleavage, and B. Whitefield for his technical contributions. We thank S. Carmeli for a sample of natural ambiguine H (**1**). We also thank D.-H. Huang and L. Pasternack for NMR spectroscopic assistance, and G. Siuzdak and R. Chadha for mass spectrometric and X-ray crystallographic assistance, respectively. We also thank Biotage for a generous donation of process vials used extensively throughout these studies. Financial support for this work was provided by The Scripps Research Institute, Amgen, AstraZeneca, the Beckman Foundation, Bristol-Myers Squibb, DuPont, Eli Lilly, GlaxoSmithKline, Pfizer, Roche, the Searle Scholarship Fund, the Sloan Foundation, NSF (predoctoral fellowship to J.M.R.) and the NIH.

Author Information The X-ray crystallographic coordinates for the structures of compounds **1** (CCDC # 623052), **2** (CCDC # 623050), and **12** (CCDC # 623051) were deposited with the Cambridge Crystallographic Data Center. These data can be obtained free of charge at www.ccdc.cam.ac.uk/retrieving.html. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.S.B. (pbaran@scripps.edu).

ARTICLES

Skew of mantle upwelling beneath the East Pacific Rise governs segmentation

Douglas R. Toomey[1], David Jousselin[2], Robert A. Dunn[3], William S. D. Wilcock[4] & R. S. Detrick[5]

Mantle upwelling is essential to the generation of new oceanic crust at mid-ocean ridges, and it is generally assumed that such upwelling is symmetric beneath active ridges. Here, however, we use seismic imaging to show that the isotropic and anisotropic structure of the mantle is rotated beneath the East Pacific Rise. The isotropic structure defines the pattern of magma delivery from the mantle to the crust. We find that the segmentation of the rise crest between transform faults correlates well with the distribution of mantle melt. The azimuth of seismic anisotropy constrains the direction of mantle flow, which is rotated nearly 10° anticlockwise from the plate-spreading direction. The mismatch between the locus of mantle melt delivery and the morphologic ridge axis results in systematic differences between areas of on-axis and off-axis melt supply. We conclude that the skew of asthenospheric upwelling and transport governs segmentation of the East Pacific Rise and variations in the intensity of ridge crest processes.

The origin of segmentation of oceanic spreading centres is controversial. According to one point of view, along-axis differences in ridge crest processes result directly from three-dimensional mantle upwelling[1–4]. Sites of vigorous volcanic and hydrothermal activity are thus thought to overlie regions of greater magma supply. Limited knowledge of mantle structure, however, has given rise to diverging opinions on the scale of three-dimensional upwellings[2–6]. Alternatively, segmentation of ridge crest processes may be regulated by the tectonic rifting of young lithosphere[7], and thus not directly linked to the form of mantle upwelling. In this case mantle flow could be either two-dimensional[8] or three-dimensional and less obvious because of efficient along-axis transport of magma by viscous flow[9].

Along the global ridge system, the fast-spreading East Pacific Rise (EPR) between the Siqueiros and Clipperton transforms (Fig. 1) currently offers our best opportunity for understanding the relations between mantle upwelling and ridge crest processes. There are several reasons for this. This section of the EPR encompasses a full spectrum of rise axis discontinuities, including: two large-offset transform faults; large, long-lived (9° 03′ N) and small, short-lived (9° 37′ N) overlapping spreading centres (OSCs)[10–12]; and smaller-scale morphologic[3], petrologic[5,12] and seismic[6,13,14] discontinuities that are typical of fast-spreading ridge segments. Accompanying these axial discontinuities are well known along-axis variations in seafloor depth, axial high morphology[15], crustal structure and thickness[16–19], lava chemistry[5,12,20,21], and seafloor hydrothermal[22,23] and biological activity[22]. These characteristics of the EPR, including the origin of ridge crest segmentation, have been hypothesized to result from the supply of magma from the mantle.

We conducted the UNDERSHOOT experiment (our data-gathering cruise) to seismically image the crustal and mantle structure between the Clipperton and Siqueiros transforms to determine the pattern of magma delivery from the mantle to the crust. Here we present the first images of mantle structure beneath an entire ridge segment bounded by long-lived tectonic discontinuities (Fig. 1). Good image resolution allows direct comparison between the scales of segmentation observed along this section of the EPR with the physical structure of the topmost mantle. Our results allow conclusions to be drawn about the driving and controlling processes for segmentation of fast-spreading ridges.

Experiment geometry and tomographic imaging

The distribution of seismic receivers and sources used to image crustal and mantle structure is shown in Fig. 1. The experiment constrains the structure of the uppermost mantle within 4 km of the Mohorovičić discontinuity and within an area extending 15 km to either side of the rise axis and 230 km along the spreading centre. A three-dimensional model of off-axis crustal structure and thickness is used to analyse the mantle refraction data (see Supplementary Information).

The P_n data (from the wave refracted below the Moho) provide good spatial sampling of mantle structure throughout the image volume (Fig. 2a). P_n travel-time residuals plotted by azimuth reveal a $\cos 2\Theta$ pattern (Fig. 2b), a signal indicative of azimuthal seismic anisotropy. The azimuth of anisotropy (that is, the fast direction for P_n propagation) is N73°E ± 1° (see Supplementary Information). Tomographic inversions, discussed below, confirm this result. The azimuth of anisotropy is rotated 9° anticlockwise with respect to the predicted spreading direction[24] (N82°E). Plotted by rise crossing point, P_n travel-time residuals show evidence for anomalously low and variable upper-mantle velocities (Fig. 2c). The average isotropic velocity that best fits the P_n data (7.6 km s^{-1}) is less than typical upper-mantle velocities, whereas delays are greater towards the centre of the transform-bounded segment and less within 20 km of the transforms.

Tomographic inversion (see Supplementary Information) of P_n travel-time data reveals a mantle low-velocity zone (MLVZ) that is segmented on a scale comparable to tectonic offsets of the EPR (Fig. 1). The MLVZ decreases in amplitude towards each transform, in agreement with the decrease in mean P_n delays (Fig. 2c). Between transforms, the MLVZ follows two *en echelon* trends that are orthogonal to the azimuth of seismic anisotropy (Fig. 1b, green lines). The *en echelon* trends are offset in a right lateral sense and rotated

[1]Department of Geological Sciences, University of Oregon, Eugene, Oregon 97403, USA. [2]Nancy-Université, CRPG, 54501 Vandoeuvre les Nancy, France. [3]Department of Geology and Geophysics, University of Hawaii-SOEST, Honolulu, Hawaii, 96822, USA. [4]School of Oceanography, University of Washington, Seattle, Washington 98195, USA. [5]Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.

anticlockwise with respect to the axis of plate spreading. Owing to this azimuthal rotation, much of the MLVZ is not centred beneath the plate boundary. Beneath and immediately north of the OSC, the MLVZ steps rightward, defining a geometrically complex transitional region that connects the *en echelon* trends of the MLVZ. The transitional region, located between 9° 00′ N and 9° 18′ N, is shifted northward of the OSC itself and coincides with a region of thicker crust[18,19].

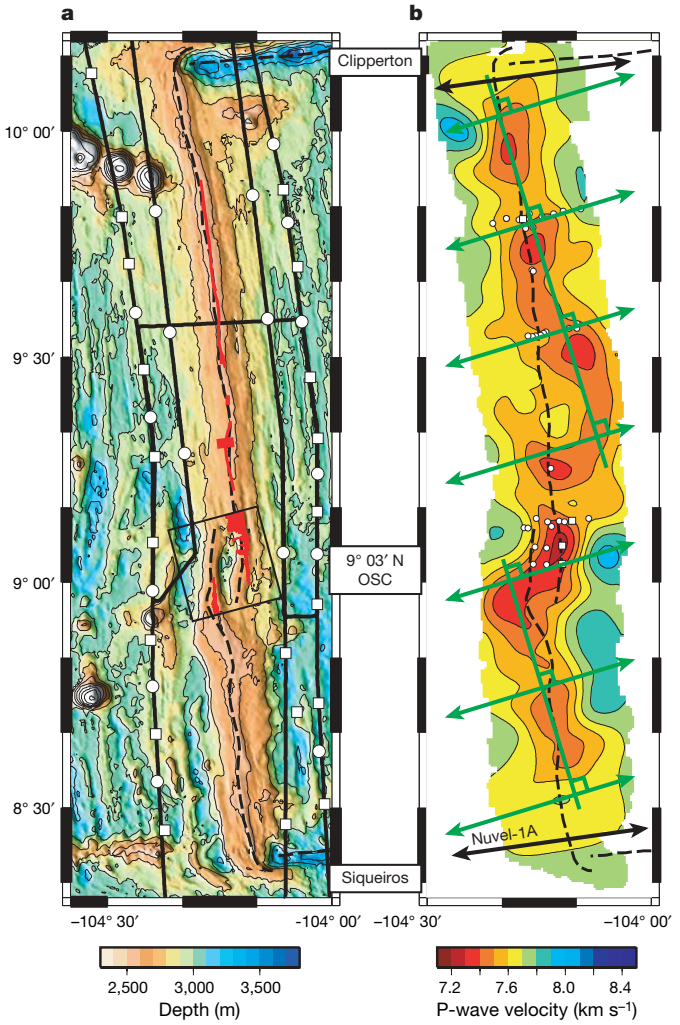


Figure 1 | Location and geometry of the seismic experiment and tomographic image of the mantle low-velocity zone (MLVZ) and orientation of mantle anisotropy. **a**, The Clipperton and Siqueiros transform faults bound the study area. Dashed lines show the location of the plate boundary. Seismic data were collected on 37 ocean-bottom receivers from the Woods Hole Oceanographic Institution. Twenty of these were Office of Naval Research, three-component seismometers (open squares) equipped with 1-Hz geophones and a hydrophone; the remaining units were ocean-bottom hydrophones (open circles). The seismic source was the RV *Maurice Ewing*'s 20-gun, 8,500-cubic-inch (139 litre) air gun array, fired at intervals of 210 s (shot spacing of 500 m) along the tracks indicated by solid black lines. Locations of axial magma chamber reflector from multichannel seismics shown in red[13,16,28]. Black box indicates the location of the three-dimensional multichannel seismics experiment. **b**, Tomographic image of mantle P-wave velocity; contour interval is 0.1 km s^{-1} and depth of section is 9 km beneath the sea floor. Green lines with arrowheads indicate azimuth of seismic anisotropy (see Figs 2b and 3); black lines with arrowheads indicate plate-spreading direction[24]. Green lines without arrowheads are perpendicular to seismic anisotropy and indicate locations of *en echelon* segments of the MLVZ. Seafloor compliance measurements are indicated by white symbols[26]; larger white squares are locations where near-Moho melt sills were detected.

North of the OSC, local minima of the MLVZ occur at intervals of approximately 25 km. Pronounced sub-axial anomalies are located near 9° 56′ N and 9° 44′ N. Equally pronounced anomalies are centred more than 10 km off-axis at 9° 32′ N and a few kilometres off-axis at 9° 15′ N. Between 9° 15′ N to 9° 35′ N the MLVZ is located 5–15 km east of the rise axis, in a region where magnetotelluric studies detect anomalously low mantle conductivities[25]. South of the OSC there is a local minima of the MLVZ near 8° 43′ N. Within our study area, the most pronounced mantle anomaly occurs beneath the eastern limb of the OSC, where several compliance measurements suggest near-Moho sills[26] and seismic imaging detects substantial amounts of melt in the lower crust[27]. The transition of the MLVZ between opposing limbs of the OSC also coincides well with a similar transition of the crustal-level axial magma chamber reflector[28] (Fig. 1).

We conducted a series of tomographic inversions for fixed models of mantle anisotropy to explore the tradeoffs between isotropic and anisotropic structure. An isotropic mantle results in a large χ-squared travel-time misfit (~6). Figure 3 shows that the data misfit is smallest

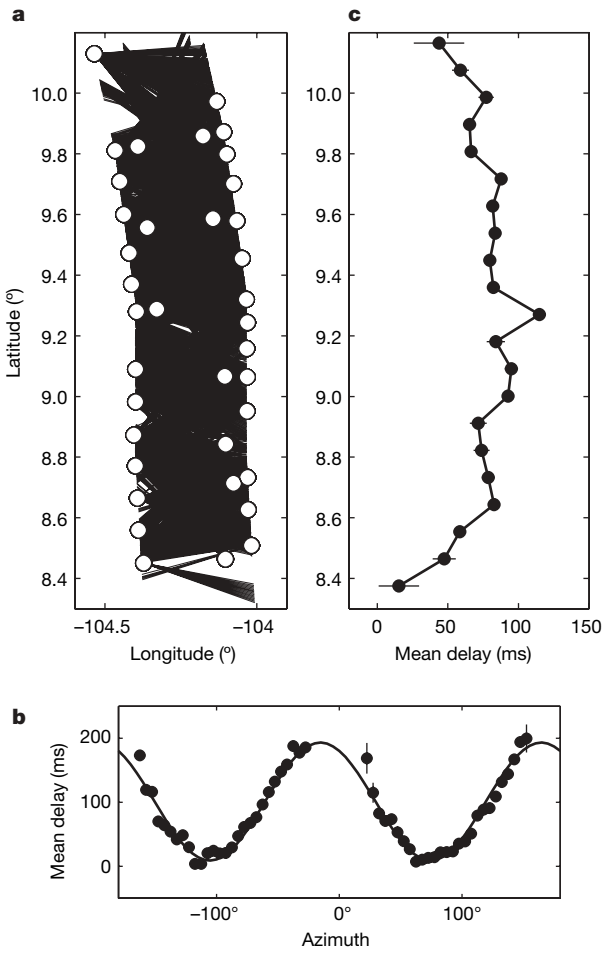


Figure 2 | Map of the distribution of seismic ray paths and mean P_n delay times versus azimuth and latitude. **a**, Map of the distribution of the 4,892 P_n ray paths used to image mantle structure. P_n data sample the mantle within about ±15 km of the rise axis. **b**, Mean P_n delay plotted by azimuth. Delays are calculated relative to an isotropic model, corrected to 40 km range and binned at intervals of 5°. Vertical bars indicate uncertainty in mean delay as determined by a Student's *t*-test (95% confidence interval). Solid line is the best-fit cos2Θ curve. The azimuth of seismic anisotropy is N73°E ± 1°. **c**, Mean P_n delay plotted by rise-axis crossing point. Delays are calculated relative to 7.8 km s^{-1}, corrected to 40 km range and have been corrected for seismic anisotropy (6% with fast axis azimuth of N73°E). Data are binned at intervals of 10 km and bars indicate uncertainty in mean delay time as determined from a Student's *t*-test (95% confidence interval).

when the azimuth of anisotropy imposed on the tomographic procedure is identical to that estimated directly from the P_n data (Fig. 2b). The data misfit is significantly larger when azimuthal anisotropy matches the predicted spreading direction. A robust result of our analysis is that both the azimuth of seismic anisotropy relative to the spreading direction and the *en echelon* trends of the MLVZ with respect to the plate boundary are rotated anticlockwise by similar amounts.

The Supplementary Information summarizes our analysis of model sensitivity and resolution and details how our current results differ from previously published studies. Therein we show that a ray-tracing error in a previous study of the 9° 03′ N OSC[29] resulted in an incorrect image.

Skew of mantle upwelling and asthenospheric flow

The isotropic component of our tomographic image constrains the distribution of melt in the topmost mantle. Regions of anomalously low seismic velocities are consistent with higher melt fractions. If melt resides in film-like geometries[30,31], then our results can be explained by the presence of 1–3% melt (see Supplementary Information); melt fractions would be greater if distributed anisotropically or if melt pockets are more spherically shaped. We attribute the MLVZ to a region of melt accumulation that lies beneath the base of newly formed oceanic crust; given the seismic wavelength, the vertical extent of this region must be several kilometres. We infer that the locus of sub-crustal melt accumulation overlies the melt production region located at depths of several tens of kilometres. The alternative, that large volumes of partially molten mantle have been transported off-axis, seems unlikely. According to this view, the axis of mantle upwelling that gives rise to decompression melting is skewed beneath the plate boundary.

The anisotropic component of our model constrains the direction of shallow mantle flow. This is because the maximum compressional wave speed for single-crystal olivine parallels the crystallographic *a* axis and deformation of mantle peridotites preferentially aligns the *a* axis in the direction of maximum shear[32,33]. Immediately beneath a spreading centre, the overturn accompanying mantle divergence generates shear strains that are many times larger than the deformation resulting from the movement of a plate over the asthenosphere[34]. We thus infer that the azimuth of seismic anisotropy beneath the spreading axis is related to the azimuth of mantle divergence. We conclude that in relation to the ridge, mantle flow is skewed by 9° with respect to the plate-spreading direction. The transport of asthenosphere away from the EPR and the axis of decompression upwelling are thus rotated anticlockwise in a coherent manner (Fig. 1).

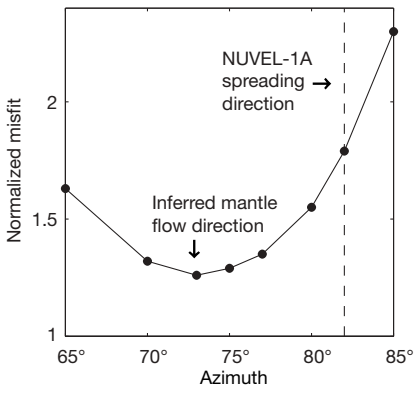


Figure 3 | Normalized data misfit following tomographic inversion versus azimuth of seismic anisotropy imposed on starting model. For each inversion, percentage (6%) and azimuth of anisotropy are held fixed. Data misfit is a minimum for an azimuth of N73°E. Misfit is significantly larger when the azimuth of anisotropy parallels the spreading direction predicted by the NUVEL-1A model[24].

The azimuthal rotation of mantle flow beneath the EPR is in the same direction as recent changes in the Euler pole for Pacific–Cocos plate motion, which has been progressing anticlockwise for the past several million years[35]. Ongoing changes in plate kinematics[35], however, are lagging behind the current direction of asthenospheric flow. We propose that the skew of mantle flow beneath the EPR is one of the driving forces for changes in plate boundary kinematics. Specifically, basal tractions imposed by mantle flow are contributing to anticlockwise changes in the spreading direction. This flow may also contribute to the transpressional and transtensional tectonics of the Clipperton and Siqueiros transforms[35], respectively (Fig. 4). The lag of the rigid plate system relative to the viscous asthenosphere may indicate that transpressive transform faults, such as the Clipperton, limit the rate at which plates adjust to changes in plate driving forces[36].

Ridge segmentation and mantle structure

Our study shows that the tectonic segmentation of the EPR correlates well with the pattern of melt delivery from the mantle to the crust. Near transforms or first-order offsets of the ridge crest[3], the increase in seismic velocities is consistent with a decrease in the amount of

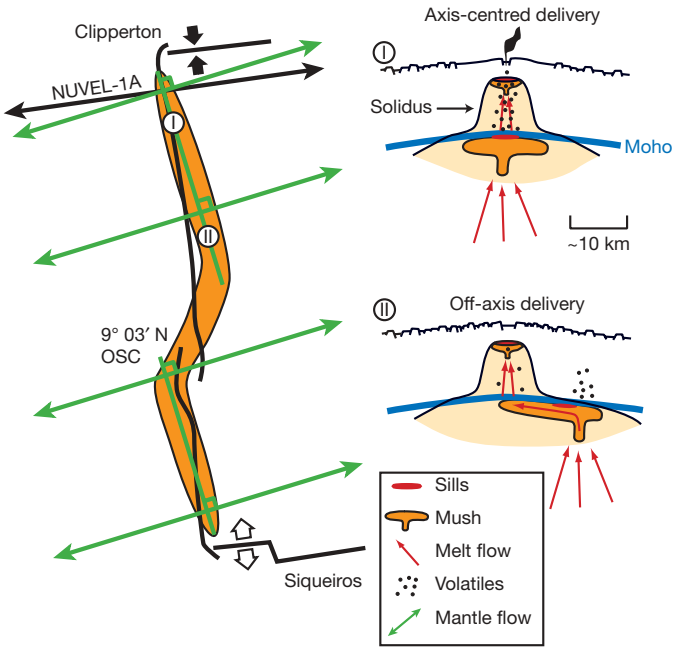


Figure 4 | Proposed model of segmentation beneath the East Pacific Rise. **a**, Map view showing plate boundary and tectonic discontinuities (solid black lines), regions of melt accumulation beneath the crust (orange area labelled 'Mush'), orientation of mantle flow (green lines with arrowheads) and trend of the *en echelon* segments of melt accumulation (green lines without arrowheads; perpendicular to mantle flow). Thin black line with arrowheads indicates NUVEL-1A spreading direction[24]. Large, paired arrows indicate regions of transpression (solid) and transtension (open) along the Clipperton and Siqueiros transform faults, respectively[35]. Circles with roman numerals indicate location of cross-sections shown in **b**. **b**, Two rise-perpendicular sections depicting magma plumbing beneath axis-centred (top) and off-axis (bottom) sites of mantle melt delivery. Above centres of melt delivery, magma accumulates at near-Moho depths, presumably within a mush-like compaction layer (orange region), with substantially lower melt concentrations in surrounding regions of the mantle. Melt migration paths indicated by red arrows; near Moho melt sills indicated by red ellipses. Small filled circles depict the exsolution of magmatic volatiles that occurs during differentiation. Sites of axis-centred delivery of mantle melt are more likely to be hydrothermally active and resurfaced frequently by extrusive volcanism (plume and smooth sea floor in top panel). Sites of off-axis delivery of mantle melt sustain greater amounts of tectonic extension, erupt less frequently and host more intermittent high-temperature hydrothermal activity (rougher sea floor in bottom panel).

melt at near-Moho depths and possibly a decrease in temperature. Towards the Siqueiros transform the increase in seismic velocity is most pronounced, consistent with the observed decrease in crustal thickness[19].

Between transform offsets, EPR segments are separated by OSCs or second-order ridge crest discontinuities[3]. The right-stepping 9° 03′ N OSC occurs near the ends of *en echelon* segments of the MLVZ, or where the centre of mantle upwelling is sheared in a right-lateral sense. We suggest that the large-scale azimuthal rotation of mantle upwelling, with respect to the current Euler pole of Pacific–Cocos plate motion, governs the formation and evolution of the OSC (Fig. 4). According to this view, mantle upwellings organize into *en echelon* segments and large-offset OSCs provide an accommodation zone that moves the lithospheric plate boundary from one segment of mantle upwelling to another. A previous hypothesis for the origin of OSCs attributes them to variations in the duration, timing or intensity of magmatic pulses that inflate crustal magma chambers and that propagate along the plate boundary, in which case a reduced magma supply is expected near second-order discontinuities[4]. An alternative model infers that OSCs are entirely lithospheric features that form in response to changes in plate motion[7]. We conclude that the segmentation of the EPR by OSCs is a direct result of *en echelon* upwelling at mantle depths. OSCs thus do not form in response to reduced magma supply, nor is their evolution related to along-axis migration of magmatic pulses away from centres of replenishment towards magma-starved discontinuities. We propose that OSCs form because the geometry of the rigid plate system is inconsistent with the underlying orientation and geometry of asthenospheric upwelling. As the pattern of mantle upwelling evolves, OSCs will respond to maintain their positions where *en echelon* segments of upwelling are sheared in the cross-axis direction. We predict that both the size and the sense of offset of an OSC, as well as its propagation history, will be related to the underlying and evolving pattern of mantle upwelling.

Between transforms and OSCs the EPR is subtly segmented at a scale of ~25 km, referred to as third-order or volcanic segmentation[3,37]. Multidisciplinary studies reveal that volcanic segments are chemically, structurally and geologically distinct. Petrologic data show that lavas erupted within volcanic segments are compositionally similar, yet differences in lava chemistry between volcanic segments can be pronounced[5,12]. Geophysical studies indicate that the centres of volcanic segments are associated with a more pronounced magmatic system at mid- to lower-crustal depths, and increased temperatures within a thermal boundary layer that connects the magmatic and hydrothermal systems[6,14,38]. Near the boundaries of volcanic segments, seismic reflection and tomographic imaging detect abrupt changes in mid- to upper-crustal structure[6,13,14,38,39]. Lastly, seafloor mapping shows that volcanic and tectonic features are also segmented at this scale[37]. The structure of the MLVZ correlates well with previous indicators of third-order segmentation, supporting the hypothesis that volcanic segmentation of the EPR is inherited from variations in mantle melt delivery[5,6,40]. We note that the spacing of volcanic segments is similar to that of diapirs mapped in the Oman ophiolite[41,42]. Additional studies will be necessary to determine whether this is a coincidence or the result of dynamic upwelling.

Ridge activity and skew of mantle upwelling

We find that the state of volcanic and tectonic activity along the ridge correlates with the cross-axis distance to a centre of mantle melt delivery. Above axis-centred delivery of mantle melt (for example, 9° 50′ N), lavas are young and fissure density is decreased[22,43,44]. Where mantle melt arrives off-axis (for example, 9° 30′ N), older lavas and increased fissuring characterize the axial high. Within our study area, the oldest lavas and the highest density of fissured sea floor[22,44] is found where the MLVZ is farthest from the rise crest (9° 20′ N). Axial eruptive styles also differ, with larger volume flows inferred where off-axis delivery of mantle melt occurs[45]. These relations suggest that shorter repose times and smaller eruption volumes may typify volcanoes fed by axis-centred delivery of mantle melt.

The vigour of hydrothermal venting also correlates with the pattern of magma delivery. In regions of axis-centred delivery, high-temperature hydrothermal venting is associated with recent volcanism and vent-supported biologic communities are abundant[22,43,44]. Either more mature vents or a lack of high-temperature vents and a general decrease in the abundance of vent-supported biota[22,43,44] are observed where mantle melt arrives off-axis. A transition from on-axis to off-axis delivery of mantle melt occurs near 9° 37′ N, where a ridge discontinuity separates a volcanic section of the axial high from a tectonic one and demarcates a significant hydrothermal boundary[12]. North of 9° 37′ N, lavas are less evolved (higher MgO content), eruption temperatures are >1,190 °C, hydrothermal venting is usually >350 °C and vent fluids are consistent with derivation from the vapour phase of phase-separated fluids[12,23]. In contrast, south of 9° 37′ N, where mantle melt is delivered off-axis, basalts are more evolved (lower MgO), eruption temperatures are <1,190 °C, hydrothermal venting temperatures are <325 °C and vent fluids are derived from the brine phase of phase-separated fluids[12,23].

Magma plumbing and the segmentation of the EPR

Three significant new findings of our study are that: (1) mantle upwelling and asthenospheric flow are skewed beneath the plate boundary, and thus the delivery of mantle melt is in many places not centred beneath the rise axis; (2) the 9° 03′ N OSC, a second-order discontinuity of the EPR, occurs where *en echelon* segments of mantle upwelling are offset in the cross-axis direction; and (3) between first- and second-order discontinuities the cross-axis offset between a centre of mantle melt delivery and the rise axis correlates with the intensity of rise crest volcanic, hydrothermal and tectonic activity. Building on these results, we propose a new model of magmatic segmentation beneath fast-spreading ridges, illustrated in Fig. 4. In our model the skew of mantle upwelling beneath the plate boundary governs second-order segmentation of the EPR and the intensity of geologic processes occurring within third-order or volcanic segments. Above we discussed the implications of our results for transforms and OSCs.

Away from OSCs and transforms, volcanic segments receive mantle-derived melt from approximately equally spaced centres. In contrast to previous hypotheses, rise-parallel variations in ridge processes are not simply a function of magma supply or the along-axis redistribution of magma away from a mantle source[3,4,9]. Instead, the cross-axis offset between a centre of mantle melt delivery and the associated axial volcano causes axis-parallel changes in ridge crest processes. We illustrate our magma-plumbing model with two cross-sections (Fig. 4). One is through an axis-centred site of mantle melt delivery and a volcano characterized by frequent extrusive volcanism and vigorous seafloor hydrothermal and biologic systems; this is representative of the rise crest near 9° 50′ N (refs 22, 43) (Figs 1 and 4). In contrast, the second cross-section is through an off-axis centre of melt delivery that feeds melt to an axial volcano where seafloor eruptions occur less frequently, surface tectonic extension is greater and seafloor hydrothermal and biologic activity are less intense; this is representative of the rise crest near 9° 30′ N (refs 22, 43, 44).

The contrasting processes at these two sites cannot be attributed to differences in the size and shape of upper crustal magma chambers, which are similar[13]. Nor is it likely that the observed differences in hydrothermal activity can be attributed to near-surface permeability, which is expected to be greater where seafloor fissuring and faulting is greater[44]. We further rule out magma supply as the controlling factor because crustal thickness, and by association the long-term magma supply, is similar at each location[18,19]. We propose that as the cross-axis offset between a centre of mantle melt delivery and the rise crest increases, so does the differentiation of magma that is delivered to the

volcano. With cooling and differentiation, magma changes its composition, increases its density and enhances the exsolution of magmatic volatiles. We infer that such fundamental changes in the qualities of magma will shape near-surface processes driven by crustal magma chambers.

Our model predicts that magma entering the crustal system from an axis-centred site of mantle melt delivery will have undergone relatively less differentiation. Such magma will be higher in temperature, more buoyant (that is, higher MgO) and retain more of its primary volatiles, thus increasing the breaching of the reaction zone above the axial magma chamber, the exchange of energy between the magmatic and hydrothermal systems and the recurrence of extrusive volcanism that localizes vigorous, high-temperature venting and biological activity. Sites of axis-centred delivery of mantle melt are most likely to be characterized as volcanically and hydrothermally robust on the basis of surface geology. In the case of off-axis delivery of mantle melt our model predicts that lateral migration of melt at sub-crustal depths will promote igneous differentiation. Crustal magma reservoirs will thus receive melt that is both cooler and denser (higher FeO) than in reservoirs above axis-centred upwellings. Depending on the degree of differentiation, considerable open-system fractionation of magmatic volatiles could occur off-axis, in which case the rate of pressurization of a ridge-crest magma chamber by volatiles would be lower. A decrease in the volatile content of an axial magma chamber should decrease the breaching of the hydrothermal reaction zone and reduce the frequency of volcanic eruptions. In this setting, localized high-temperature hydrothermal venting may be intermittent and lower-temperature, diffuse-flow venting could prevail. Such characteristics are commonly attributed to magma starvation. In our magma plumbing model, two axial volcanoes (or third-order segments) can receive similar volumes of magma, but behave differently (extrusive versus intrusive) because of sub-crustal, differentiation-induced changes in magma quality.

Previous models of mid-ocean ridges have usually assumed that magma supply controls segmentation, that asthenospheric transport parallels the spreading direction and that mantle upwelling and melt delivery is symmetric about the rise axis. Ours is the first study to show large-scale skew of mantle upwelling beneath mid-ocean ridges and as such it renews the debate over the origin and significance of spreading-centre segmentation. One implication of our results is that local plate motions alone are not the sole cause of sub-ridge mantle flow. On the contrary, the skew of mantle upwelling and transport can act as a driving force for the tectonic reorganization of the EPR and cause along-axis variations in the intensity of ridge-crest processes. We speculate that the skew of the sub-ridge asthenosphere owes its origin to global patterns of mantle flow, which are strongly influenced by the viscous coupling between subducting oceanic slabs and the surrounding mantle. If this speculation holds, it implies that the flux of slabs into the mantle may be linked to the segmentation of mid-ocean ridges.

Received 30 August 2006; accepted 8 February 2007.

1. Whitehead, J. A. Jr, Dick, H. J. B. & Schouten, H. A mechanism for magmatic accretion under spreading centres. *Nature* **312,** 146–148 (1984).
2. Schouten, H., Klitgord, K. D. & Whitehead, J. A. Segmentation of mid-ocean ridges. *Nature* **317,** 225–229 (1985).
3. Macdonald, K. C. *et al.* A new view of the mid-ocean ridge from the behaviour of ridge-axis discontinuities. *Nature* **335,** 217–225 (1988).
4. Macdonald, K. C., Scheirer, D. S. & Carbotte, S. M. Mid-ocean ridges: Discontinuities, segments and giant cracks. *Science* **253,** 986–994 (1991).
5. Langmuir, C. H., Bender, J. F. & Batiza, R. Petrological and tectonic segmentation of the East Pacific Rise, 5°30′-14°30′N. *Nature* **322,** 422–429 (1986).
6. Toomey, D. R., Purdy, G. M., Solomon, S. C. & Wilcock, W. S. D. The three-dimensional seismic velocity structure of the East Pacific Rise near latitude 9°30′N. *Nature* **347,** 639–645 (1990).
7. Lonsdale, P. Segmentation of the Pacific-Nazca Spreading Center, 1°N-20°S. *J. Geophys. Res.* **94,** 12197–12226 (1989).
8. Parmentier, E. M. & Morgan, J. P. Spreading rate dependence of three-dimensional structure in oceanic spreading centers. *Nature* **348,** 325–328 (1990).
9. Bell, R. E. & Buck, W. R. Crustal control of ridge segmentation inferred from observations of the Reykjanes ridge. *Nature* **357,** 583–586 (1992).
10. Sempéré, J.-C. & Macdonald, K. C. Deep-tow studies of the overlapping spreading centers at 9°03′N on the East Pacific Rise. *Tectonics* **5,** 881–900 (1986).
11. Carbotte, S. M. & Macdonald, K. C. East Pacific Rise 8°-10°30′N: Evolution of ridge segments and discontinuities from SeaMARC II and three-dimensional magnetic studies. *J. Geophys. Res.* **97,** 6959–6982 (1992).
12. Smith, M. C. *et al.* Magmatic processes and segmentation at a fast spreading mid-ocean ridge; detailed investigation of an axial discontinuity on the East Pacific Rise crest at 9°37′N. *Geochem. Geophys. Geosyst.* **2,** doi:10.1029/2000GC000134 (2001).
13. Kent, G. M., Harding, A. J. & Orcutt, J. A. Distribution of magma beneath the East Pacific Rise between the Clipperton Transform and the 9°17′N Deval from forward modeling of common depth point data. *J. Geophys. Res.* **98,** 13945–13969 (1993).
14. Dunn, R. A., Toomey, D. R. & Solomon, S. C. Three-dimensional seismic structure and physical properties of the crust and shallow mantle beneath the East Pacific Rise at 9°30′N. *J. Geophys. Res.* **105,** 23537–23555 (2000).
15. Scheirer, D. S. & Macdonald, K. C. Variation in cross-sectional area of the axial ridge along the East Pacific Rise: Evidence for the magmatic budget of a fast spreading center. *J. Geophys. Res.* **98,** 7871–7885 (1993).
16. Detrick, R. S. *et al.* Multi-channel seismic imaging of a crustal magma chamber along the East Pacific Rise. *Nature* **326,** 35–41 (1987).
17. Vera, E. E. *et al.* The structure of 0- to 0.2-m.y.-old oceanic crust at 9°N on the East Pacific Rise from expanded spread profiles. *J. Geophys. Res.* **95,** 15529–15556 (1990).
18. Barth, G. A. & Mutter, J. C. Variability in oceanic crustal thickness and structure: Multichannel seismic reflection results from the northern East Pacific Rise. *J. Geophys. Res.* **101,** 17951–17975 (1996).
19. Canales, J. P., Detrick, R. S., Toomey, D. R. & Wilcock, W. S. D. Segment-scale variations in the crustal structure of 150–300 kyr old fast spreading oceanic crust (East Pacific Rise, 8°15′N-10°5′N) from wide-angle seismic refraction profiles. *Geophys. J. Int.* **152,** 766–794 (2003).
20. Batiza, R. & Niu, Y. Petrology and magma chamber processes at the East Pacific Rise ~9°30′N. *J. Geophys. Res.* **97,** 6779–6797 (1992).
21. Perfit, M. R. *et al.* Small-scale spatial and temporal variations in mid-ocean ridge crest magmatic processes. *Geology* **22,** 375–379 (1994).
22. Haymon, R. M. *et al.* Hydrothermal vent distribution along the East Pacific Rise crest (9°09′-54′N) and its relationship to magmatic and tectonic processes on fast-spreading mid-ocean ridges. *Earth Planet. Sci. Lett.* **102,** 513–534 (1991).
23. Von Damm, K. L. Chemistry of hydrothermal vent fluids from 9°-10°, East Pacific Rise: ''Time zero,'' the intermediate posteruptive period. *J. Geophys. Res.* **105,** 11203–11222 (2000).
24. Gripp, A. E. & Gordan, R. G. Young tracks of hotspots and current plate velocities. *Geophys. J. Int.* **150,** 321–361 (2002).
25. Key, K. & Constable, S. Mantle upwelling beneath the East Pacific Rise at 9°30′N. *Eos* (Fall Meet. Suppl.) **87** (52), abstr. B31B–1114 (2006).
26. Crawford, W. C. & Webb, S. C. Variations in the distribution of magma in the lower crust and at the Moho beneath the East Pacific Rise at 9°-10°N. *Earth Planet. Sci. Lett.* **203,** 117–130 (2002).
27. Singh, S. C. *et al.* Seismic reflection images of the Moho underlying melt sills at the East Pacific Rise. *Nature* **442,** 287–290 (2006).
28. Kent, G. M. *et al.* Evidence from three-dimensional seismic reflectivity images for enhanced melt supply beneath mid-ocean-ridge discontinuities. *Nature* **406,** 614–618 (2000).
29. Dunn, R. A., Toomey, D. R., Detrick, R. S. & Wilcock, W. S. D. Continuous mantle melt supply beneath an overlapping spreading center on the East Pacific Rise. *Science* **291,** 1955–1958 (2001).
30. Faul, U. H., Toomey, D. R. & Waff, H. S. Intergranular basaltic melt is distributed in thin, elongated inclusions. *Geophys. Res. Lett.* **21,** 29–32 (1994).
31. Hammond, W. C. & Humphreys, E. D. Upper mantle seismic wave velocity: Effects of realistic partial melt geometries. *J. Geophys. Res.* **105,** 10975–10986 (2000).
32. Nicolas, A. & Christensen, N. I. in *Composition, Structure, and Dynamics of the Lithosphere-Asthenosphere System* (eds Fuchs, K. & Froidevaux, C.) 111–123 (American Geophysical Union, Washington DC, 1987).
33. Ben Ismaïl, W. & Mainprice, D. An olivine fabric database: an overview of upper mantle fabrics and seismic anisotropy. *Tectonophysics* **296,** 145–157 (1998).
34. Blackman, D. K., Wenk, H.-R. & Kendall, J. M. Seismic anisotropy of the upper mantle: 1. Factors that affect mineral texture and effective elastic properties. *Geochem. Geophys. Geosyst.* **3,** doi:10.1029/2001GC000248 (2002).
35. Pockalny, R. A., Fox, P. J., Fornari, D. J., Macdonald, K. C. & Perfit, M. R. Tectonic reconstruction of the Clipperton and Siqueiros Fracture Zones: Evidence and consequences of plate motion change for the last 3 Myr. *J. Geophys. Res.* **102,** 3167–3181 (1997).
36. Richards, M. A. & Lithgow-Bertelloni, C. Plate motion changes, the Hawaiian-Emperor bend, and the apparent success and failure of geodynamic models. *Earth Planet. Sci. Lett.* **137,** 19–27 (1996).
37. White, S. M., Haymon, R. M., Fornari, D. J., Perfit, M. R. & Macdonald, K. C. Correlation between volcanic and tectonic segmentation of fast-spreading ridges: Evidence from volcanic structures and lava flow morphology on the East Pacific Rise at 9°-10°N. *J. Geophys. Res.* **107,** doi:10.1029/2001JB000571 (2002).
38. Toomey, D. R., Solomon, S. C. & Purdy, G. M. Tomographic imaging of the shallow crustal structure of the East Pacific Rise at 9°30′N. *J. Geophys. Res.* **99,** 24135–24157 (1994).

39. Tian, T., Wilcock, W. S. D., Toomey, D. R. & Detrick, R. S. Seismic heterogeneity in the upper crust near the 1991 eruption site on the East Pacific Rise. *Geophys. Res. Lett.* **27,** 2369–2372 (2000).
40. Dunn, R. A. & Toomey, D. R. Seismological evidence for three-dimensional melt migration beneath the East Pacific Rise. *Nature* **388,** 259–262 (1997).
41. Nicolas, A. *Structures of Ophiolites and Dynamics of Oceanic Lithosphere* 70–77 (ed. Nicolas, A.) (Kluwer Academic, Dordrecht, 1989).
42. Jousselin, D., Nicolas, A. & Boudier, F. Detailed mapping of a mantle diapir below a paleo-spreading center in the Oman ophiolite. *J. Geophys. Res.* **103,** 18153–18170 (1998).
43. Fornari, D. J., Haymon, R. M., Perfit, M. R., Gregg, T. K. P. & Edwards, M. H. Axial summit trough of the East Pacific Rise 9°-10°N: Geological constraints and evolution of the axial zone of fast spreading mid-ocean ridges. *J. Geophys. Res.* **103,** 9827–9855 (1998).
44. Wright, D. J., Haymon, R. M. & Fornari, D. J. Crustal fissuring and its relationship to magmatic and hydrothermal processes on the East Pacific Rise crest (9°12′ to 54′N). *J. Geophys. Res.* **100,** 6097–6120 (1995).
45. Soule, S. A. *et al.* Channelized lava flows at the East Pacific Rise crest 9°-10°N: The importance of off-axis lava transport in developing the architecture of young oceanic crust. *Geochem. Geophys. Geosyst.* **6,** doi:10.1029/2005GC000912 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the officers and crew of the RV *Maurice Ewing* and members of the scientific party for their assistance. D.R.T. thanks E. Hooft and E. Humphreys for numerous discussions and J. Karson, T. Durant and D. Villagomez for comments. Supported by the RIDGE and RIDGE 2000 Programs, Ocean Sciences Division, NSF.

Author Contributions All authors participated in the experimental design, the collection of the data and in several stages of data reduction and analysis. D.R.T. conducted the tomographic analysis and wrote the manuscript with comments from co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.R.T. (drt@uoregon.edu).

ARTICLES

Autocatalytic cleavage of *Clostridium difficile* toxin B

Jessica Reineke[1]*, Stefan Tenzer[2]*, Maja Rupnik[3], Andreas Koschinski[4], Oliver Hasselmayer[1], André Schrattenholz[5], Hansjörg Schild[2] & Christoph von Eichel-Streiber[1]

***Clostridium difficile*, the causative agent of nosocomial antibiotic-associated diarrhoea and pseudomembranous colitis, possesses two main virulence factors: the large clostridial cytotoxins A and B. It has been proposed that toxin B is cleaved by a cytosolic factor of the eukaryotic target cell during its cellular uptake. Here we report that cleavage of not only toxin B, but also all other large clostridial cytotoxins, is an autocatalytic process dependent on host cytosolic inositolphosphate cofactors. A covalent inhibitor of aspartate proteases, 1,2-epoxy-3-(*p*-nitrophenoxy)propane, completely blocked toxin B function on cultured cells and was used to identify its catalytically active protease site. To our knowledge this is the first report on a bacterial toxin that uses eukaryotic signals for induced autoproteolysis to deliver its toxic domain into the cytosol of target cells. On the basis of our data, we present an integrated model for the uptake and inositolphosphate-induced activation of toxin B.**

With the introduction of antibiotic therapy into clinical practice the lifespan of the population in the industrialized world increased and infections seemed to become manageable. However, this observation is deceptive, because many microbes possess or develop resistance mechanisms[1,2]. Furthermore, antibiotics may cause prominent side effects, like the observed increase of pseudomembranous colitis when clindamycin was introduced in the 1970s[3,4]. Today, we know that nearly all known antibiotics are capable of inducing antibiotic-associated diarrhoea or its fatal form, pseudomembranous colitis[5,6].

The factors inducing antibiotic-associated diarrhoea or pseudomembranous colitis are two toxins produced by *Clostridium difficile*, toxin A (TcdA) and toxin B (TcdB)[7,8], which are homologous to each other[9], to TcsH and TcsL of *Clostridium sordellii*[10] and to Tcnα of *Clostridium novyi*[11]. Because of their sequence homology, similar domain structure and glycosyltransferase properties these toxins are designated 'large clostridial cytotoxins' (LCTs)[12,13].

Recently, *C. difficile* attracted great attention[14,15] because hypervirulent, ribotype 027 isolates, which are resistant to fluoroquinolones, were reported in Canada and the USA. These reports and some recent data of younger people encountering severe *C. difficile* disease without prior antibiotic treatment indicate a changing epidemiology and call for a shift in the perception that only elderly, immunocompromised people are at high risk of suffering from *C. difficile* infection, mainly following hospitalization[16]. There is the threat that such strains will spread worldwide and thus compromise antibiotic treatment in general[17,18]. Accordingly, a better understanding of the mode of action of TcdA and TcdB is urgently needed.

TcdA and TcdB are single-chained and have a tripartite functional organization[12]: the carboxy-terminal domain is responsible for binding to the host cell membrane[19]; the hydrophobic middle part is supposed to be involved in translocation across cellular membranes[20]; and the amino-terminal domain harbours the catalytically active glucosyltransferase site[12,21]. Although knowledge about the uptake process is still fragmentary, it is generally accepted that on binding to host cell receptors[22,23] the toxins are endocytosed[24]. The translocation of the toxins to the cytosol occurs from early endosomal compartments[25,26]. Because TcdA forms pores in artificial membranes at low pH (ref. 27), it was suggested that in living cells the low endosomal pH triggers similar conformational changes and accordingly may propagate translocation via pore formation.

From early on, it was postulated that the toxin is activated through catalytic processing[26,28]. Recently, it was shown that TcdB is cleaved between Leu 543 and Gly 544 of TcdB at neutral pH (ref. 29), generating fragments of 63 kDa (catalytic-DXD domain) and 207 kDa (ligand-translocation domain). Only the smaller N-terminal fragment harbouring the glucosyltransferase activity reaches the cytosol, whereas the 207 kDa C-terminal fragment remains in the membrane fraction[20,29]. Microinjection experiments showed that the 63 kDa fragment exhibits full cytotoxic activity shortly after injection, whereas the holotoxin starts with low activity, which increases with time[29], indicating the biologic relevance of proteolytic processing for catalytic activity. The released cytotoxic fragment can then act on its cytosolic targets—the GTPases of the Rho/Rac family[30,31]. Until now, the assumption was that this cleavage is catalysed by a pepstatin A-sensitive host cell protease[20,29].

Results

Identification of the host factor. To identify the host factor required for cleavage of TcdB, we used TcdB preparations[32] that were Cy3-labelled as described in the literature[29]. Cytosolic extracts of porcine splenocytes showed strong cleavage activity (Fig. 1a) and thus served as raw material for further analysis. Purification of cytosolic fractions of splenocyte extracts (Supplementary Fig. 1) led to the following inconsistencies with the thesis that we had purified a cellular protease: concentration by ammonium sulphate precipitation was not possible, activity was fully retained in heated extracts (96 °C, 30 min), effective TcdB cleavage was induced with the aqueous protein-free phase of phenol chloroform extractions and highly purified active

[1]Johannes-Gutenberg Universität Mainz, Institut für medizinische Mikrobiologie and Hygiene, Hochhaus am Augustusplatz, 55131 Mainz, Germany. [2]Johannes-Gutenberg Universität Mainz, Institut für Immunologie, Hochhaus am Augustusplatz, 55131 Mainz, Germany. [3]University of Maribor, Faculty of Medicine, Slomskov trg 15 and Institute of Public Health Maribor, Prvomajska 1, 2000 Maribor, Slovenia. [4]Justus-Liebig Universität Giessen, Rudolf-Buchheim-Institut für Pharmakologie, Frankfurter Strasse 107, 35392 Giessen, Germany. [5]ProteoSys AG, Carl-Zeiss-Strasse 51, 55129 Mainz, Germany.
*These authors contributed equally to this work.

extracts showed no significant absorption at 280 nm (Supplementary Fig. 1).

The active components were identified by nano-electrospray-ionization quadrupole time-of-flight (nanoESI-QTOF) mass spectrometry of purified active fractions (Fig. 1b). The comparison of fragment spectra of the active substance (Fig. 1c, d) with published spectra[33] revealed strong similarities to inositolphosphates (Fig. 1e). Several inositolphosphates (InsPs) were tested in the cleavage assay and inositolhexaphosphate ($InsP_6$) showed the highest cleavage activity (see Supplementary Table). Unlabelled TcdB visualized by zinc staining[34] showed activity of $InsP_6$ down to 1 μM (Supplementary Fig. 2). Because the total cellular $InsP_6$ concentration is in the range of 15–100 μM (ref. 35), concentrations of $InsP_6$ required for *in vitro* cleavage can be regarded as physiologically relevant. No cleavage activity has ever been observed or attributed to InsPs itself. Whereas activation of proteins by phosphorylation is a well known mechanism, inositolphosphates, in contrast to inositolpyrophosphates[36], do not contain high energy pyrophosphate bonds. It is therefore unlikely that inositolphosphates activate the toxin by phosphorylation.

Exclusion of contaminant proteases. In accordance with previous publications[29], cleavage was pepstatin A-sensitive, indicating that the TcdB preparation contained a latent, $InsP_x$-inducible aspartate protease (data not shown). A series of experiments was performed to exclude the contribution of a contaminating $InsP_x$-inducible protease that might co-purify with TcdB. A putative unspecific action of our TcdB preparation on other proteins was excluded (Supplementary Fig. 3). Using monoclonal antibody 2CV (ref. 37), TcdB was additionally affinity purified, which yielded more than 98% pure TcdB as assessed by SDS–polyacrylamide gel electrophoresis (PAGE). The remaining contaminant proteins were identified by tandem mass spectrometry (MS/MS) analysis. All proteins matched to sequences from the recently published *C. difficile* genome[38] and showed no similarity to known endoproteases (data not shown). The affinity-purified TcdB retained the $InsP_6$-inducible cleavage activity. Further, cleavage of TcdB by a contaminant protease would be an intermolecular process, whereas autocatalysis is intramolecular. Because dilution favours intramolecular reactions in comparison to intermolecular reactions, we performed cleavage of TcdB at different concentrations and observed that in contrast to the partial cleavage at 5 μM TcdB, cleavage is nearly completed after 30 min at a concentration of 50 nM TcdB (Fig. 2a). This result is inconsistent with the participation of a contaminating protease, which would show reduced activity at lower concentrations. Consequently, we postulated autocatalytic cleavage by an intrinsic TcdB aspartate protease.

Proof of autocatalytic TcdB cleavage. To prove further our model for autocatalytic TcdB activation, we decided to inactivate and identify the

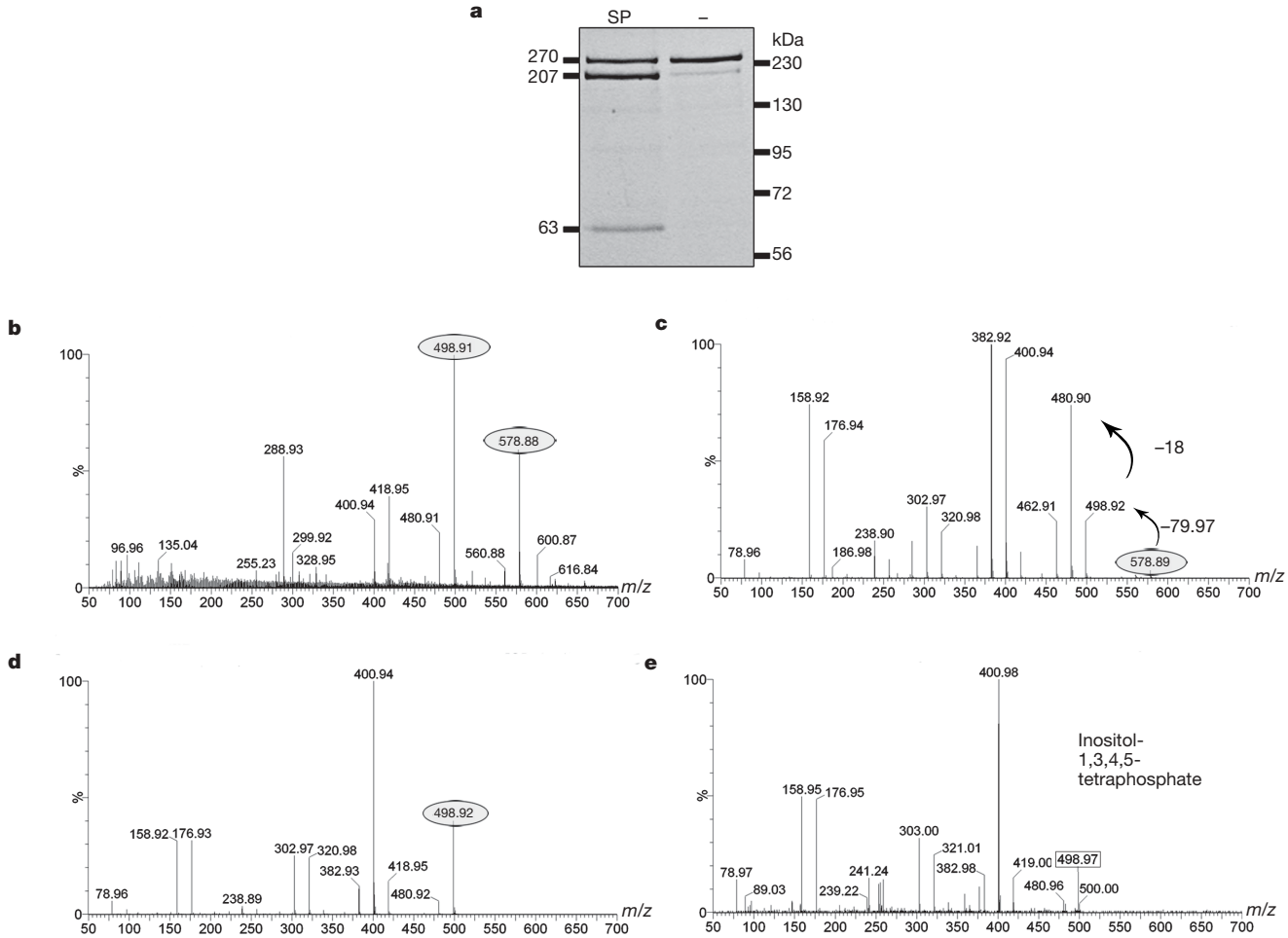


Figure 1 | Purified cytosolic splenocyte extract in the *in vitro* TcdB cleavage assay and mass spectrometry. **a**, *In vitro* cleavage reaction. Cleavage of fluorescently labelled Cy3-TcdB activated by splenocyte cytosolic extract (SP) and the same reaction without SP (−); visualization was by Cy3 fluorescence. **b**–**e**, Active fractions of the splenocyte extract purification (Supplementary Fig. 1) were analysed by mass spectrometry; representative spectra are shown. $InsP_x$-peaks from the purified porcine splenocyte extract are encircled. The *y* axis measures MS-intensity as a per cent of the most intense peak. *m*/*z*, mass-to-charge ratio. **b**, Survey spectra from fraction C5 of S75 size exclusion chromatography (Supplementary Fig. 1f) were acquired over a mass range from 50–1,500 Da. **c**, **d**, From fraction C5 the peaks at 578.88/578.89 Da (encircled and shaded) and 498.91/498.92 Da (encircled and shaded) were manually selected for fragmentation in MS/MS mode (shown in **c** and **d**, respectively). **e**, The MS/MS fragment spectrum of commercially available inositol-1,3,4,5-tetraphosphate (boxed, with a peak at 498.97) served as a positive control.

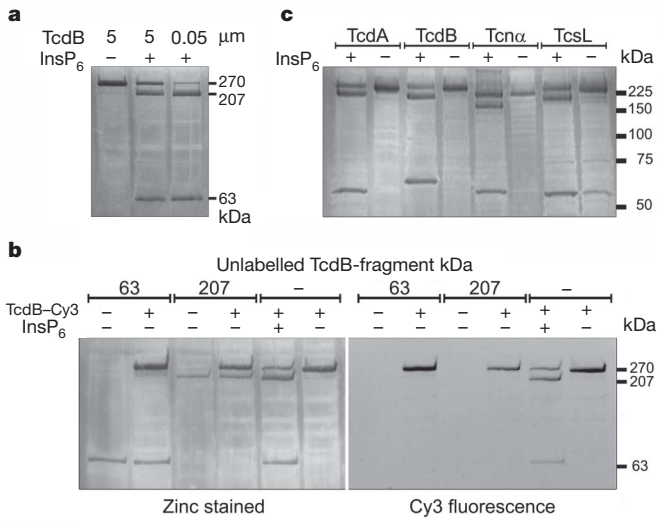


Figure 2 | *In vitro* cleavage of LCTs by purified splenocyte extracts or inositolhexaphosphate. **a**, Cleavage of native TcdB induced by the addition of $InsP_6$. Unlabelled toxin (16 µg) was incubated with $InsP_6$ (at 100 µM concentration) at different TcdB concentrations (5.0 or 0.05 µM); following cleavage, the solutions were lyophilized, resuspended in water (approximately 1 µg) subjected to SDS–PAGE and detected with zinc staining. **b**, Exclusion of a cascade reaction. Cy3-labelled TcdB was co-incubated with the isolated (pre-activated) 63 kDa catalytic-DXD or the 207 kDa ligand-translocation domain, respectively. Left panel, zinc staining; right panel, the same gel visualized by Cy3 fluorescence. **c**, Cleavage of LCTs from different *Clostridia*. Assay without (−) or with (+) activation by 10 mM $InsP_6$. LCTs were: TcdA and TcdB of *C. difficile* strain VPI10463; Tcnα of *C. novyi* strain 19402; and TcsL of *C. sordellii* strain IP-82. Visualization was by zinc staining.

active centre of the protease. EPNP (1,2-epoxy-3-(*p*-nitrophenoxy)propane) is a known inhibitor of aspartate proteases, which, in contrast to pepstatin A, covalently binds to the active site[39]. To determine the TcdB intrinsic protease active site we thus generated an 'EPNP-derivative' of TcdB. Following addition of $InsP_6$, this modified TcdB was no longer cleaved (data not shown). This suggests that TcdB had lost its aspartate protease function. The active site of aspartate proteases is known to be situated in the motif DXG (where X is S or T)[39]. According to database searches (see Supplementary Information), TcdB contains three of these motifs at positions 1665, 2001 and 2133. To exclude the possibility that EPNP modified a contaminating protease rather than TcdB, and to identify its active site, we analysed fragments generated by tryptic proteolysis of EPNP-treated and untreated TcdB by nanoESI-QTOF mass spectrometry. Only a single fragment with an EPNP-induced mass shift of 195 Da was detected. The triple charged fragment had an apparent mass of $m/z = 1075.49$ and was confirmed by MS/MS analysis to be amino acids 1652–1678 of TcdB (Supplementary Fig. 4). This fragment contains the aspartate protease consensus motif DSG at amino acid position 1665, confirming the predicted DSG(1665) to be the active site of the autocatalytic TcdB protease. Furthermore, EPNP pretreatment completely abrogated cellular toxicity when the toxin was added to the target cell culture (Fig. 3c). As expected, loss of cytotoxicity was not due to modification of the catalytic-DXD glucosyltransferase domain, because microinjection of EPNP-treated TcdB induced cell rounding (Fig. 3f).

To dissect further the autocatalytic cleavage mechanism, we isolated both N- and C-terminal fragments by anion exchange chromatography and tested their proteolytic activity. As indicated in Fig. 2b, neither the 63 kDa nor the 207 kDa fragment was able to cleave full-length Cy3-labelled TcdB in the absence of $InsP_6$.

Thus the cleavage reaction of TcdB is intramolecular and not followed by a cascade reaction catalysed by its cleavage fragments.

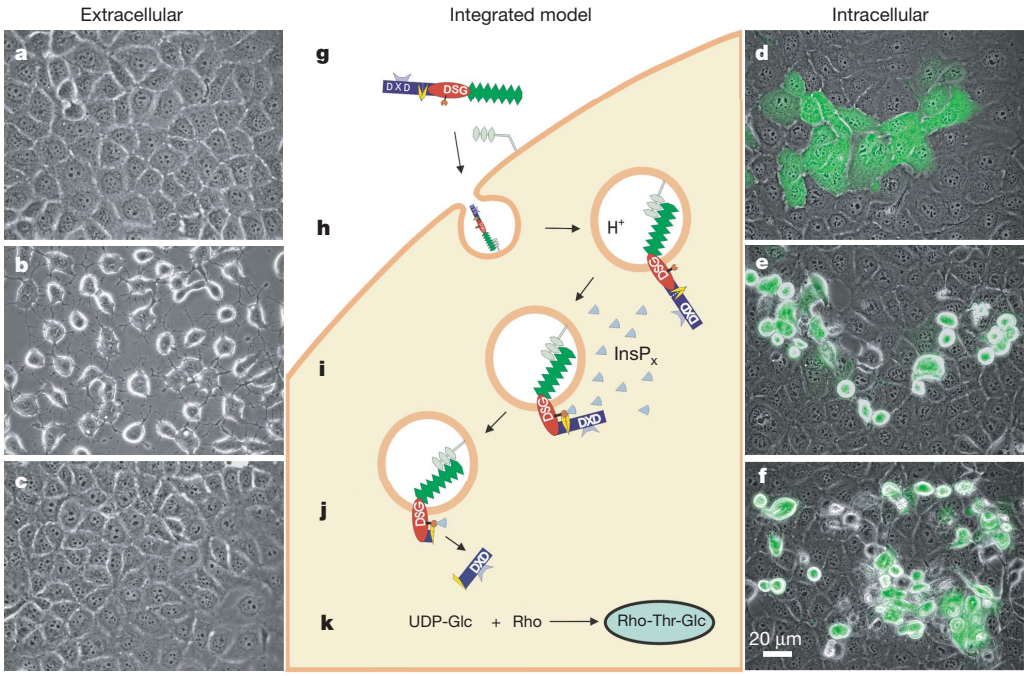


Figure 3 | An integrated model of events necessary for the cellular cytotoxic action of TcdB. **a**–**c**, Cytotoxic effects of TcdB (1 µg ml^{-1}) applied to the supernatant of IHKE cells. **a**, Control buffer without TcdB. **b**, Unmodified TcdB. **c**, EPNP-modified TcdB. **d**–**f**, Effects of TcdB microinjected into the cytosol of IHKE cells. **d**, Control buffer (containing dextran-FITC) without TcdB. **e**, Unmodified TcdB. **f**, EPNP-modified TcdB. **g**–**k**, Schematic model of TcdB uptake and delivery of the catalytic-DXD domain into the eukaryotic cell. Colour code: dark green, the TcdB ligand domain; red, the TcdB translocation domain; dark blue, the TcdB catalytic-DXD domain; light grey-green, the receptor for TcdB; brown, cellular membranes; and light blue, cytosolic inositolphosphates. **g**, Interaction with cell receptors induces receptor-mediated endocytosis. **h**, The acidic pH of the endosome triggers the first conformational change and results in pore formation of the ligand-translocation domain. **i**, Cytosolic $InsP_x$ interacts with TcdB and induces a second conformational change, activating the intrinsic protease function. **j**, Autocatalytic cleavage and release of the catalytic-DXD glucosyltransferase domain into the cytosol. **k**, Glucosylation of the cytosolic target GTPases, leading to disaggregation of the cytoskeleton and cell death.

Furthermore, Fig. 2b shows that TcdB loses its intrinsic protease function following cleavage.

Autocatalytic cleavage of other LCTs. Assuming autocatalytic activation, we expected that other members of the LCT family[12,40] would show similar properties. An assay with $InsP_6$-activated *C. sordellii* TcsL, *C. novyi* Tcnα and *C. difficile* TcdA demonstrated cleavage like that observed with TcdB (Fig. 2c). Apparently, $InsP_x$-induced autocatalytic cleavage is a common property of all LCTs.

Model for the uptake and activation of toxin B. A model of the events occurring within the cells is presented in Fig. 3. On binding to host cell receptors[22,23] (Fig. 3g) the toxins are endocytosed[24]. The low pH in endosomes triggers the first conformational change of the toxin leading to membrane insertion[26,28,41] (Fig. 3h), followed by translocation of the N-terminal catalytic-DXD domain into the cytosol[20,29].

The translocation-ligand domain remains associated with endosomal membranes and only the catalytic-DXD domain penetrates into the cytosol[20,29]. We show here that cytosolic $InsP_x$ is needed to induce autocatalytic cleavage of TcdB. It is known that in eukaryotic cells InsPs may induce conformational changes of proteins[42,43], and therefore we postulate that TcdB undergoes a second conformational change (Fig. 3i), which is induced by $InsP_x$ and immediately leads to autocatalytic proteolysis and delivery of the catalytic-DXD domain into the cytosol (Fig. 3j). Because our experimental data show that the intramolecular reaction quickly follows the addition of $InsP_6$, cleavage of the glucosyltransferase portion could immediately occur on its translocation into the cytosol. Additionally, our data using EPNP-modified TcdB show that $InsP_x$-induced autocatalytic cleavage is absolutely required for the execution of TcdB's biological function. Once the toxin's aspartate protease function is blocked (Fig. 3c), the catalytic-DXD domain can no longer be cleaved off and thus, according to other researchers[20,29], the DXD domain can no longer reach the cytosol. Following its translocation and release, the catalytic-DXD fragment acts on its cytosolic targets (Fig. 3k), the GTPases of the Rho/Rac family, leading to the observed blockade of signal transduction processes and, consequently, the disaggregation of the cytoskeleton and cell death[12,31,40].

Discussion

In conclusion, we identified $InsP_x$ as the host factor required for the autocatalytic proteolysis of TcdB. Cellular inositolphosphates are involved in the regulation of a large variety of cellular processes in the cytoplasm, the plasma membrane and the nucleus[43,44]. Recently, it was reported that $InsP_6$ is required for the correct folding of ADAR2, involved in RNA editing[42]. However, $InsP_x$ as a cofactor for activation of proteolytic activity has not been reported before, to the best of our knowledge. Furthermore, our data and some recent reports[42,43] indicate that in eukaryotic cells InsPs serve more generally as cofactors for proper protein folding and function. The proof of $InsP_6$-induced cleavage of LCTs shown in Fig. 2c leads us to propose $InsP_x$-inducible proteolytic cleavage as an additional feature common to the LCT family. However, in contrast to TcdA, TcdB and TcsL, which have a DXG-motif at the same position, Tcnα has no DXG-site and therefore should possess a different protease function to the aspartate protease of the other LCTs.

Regarding the similar genomic organization of LCTs[45], it is not surprising that they carry their intramolecular protease with them because this combines complete mechanisms of regulated production, release of toxins and induced self-activation at the final site of action[46]. During evolution, LCT-producing *Clostridia* have thus gained a very effective means of controlling the energy-demanding use of their toxins.

METHODS

Labelling of toxins with Cy3. The toxins TcdB and TcdA produced by the toxin high-producer strain VPI10463 were purchased from tgcBIOMICS (Mainz, Germany). Aliquots of 200–400 μg were transferred to $NaHCO_3$ buffer by gel filtration and labelled with Cy3 (GE Healthcare) according to the manufacturer's instructions. Unbound dye was removed by size exclusion chromatography using 10 mM Tris-HCl, pH 8.5, as running buffer. The molar dye/toxin-labelling ratio amounted to 0.8–1.6. Labelled toxins were aliquoted and stored at −80 °C.

***In vitro* cleavage assay.** The cleavage assay was performed as described previously[29], in Tris-HCl, pH 8.5, buffer at a volume of 10 μl (unless stated otherwise) consisting of 500 ng of labelled toxin or 1 μg of unlabelled toxin, 3 μl of cell extract and purified fractions or commercially available inositol phosphates (purchased from Axxora, Merck, and Sigma, respectively). Following a 30 min incubation at 25 °C, samples were separated by SDS–PAGE (10%), Cy3-labelled toxin and toxin fragments were visualized by fluorescence, unlabelled toxins were visualized by zinc staining and both were documented with a Gel Doc EQ System Image Reader (Bio-Rad).

Preparation of cytosolic porcine splenocyte extract. Freshly isolated pig spleens were prepared in phosphate buffered saline (PBS) and processed into single-cell suspensions. Erythrocytes were lysed by adding low-salt buffer (150 mM NH_4Cl, 1 mM $KHCO_3$ and 0.1 mM EDTA, pH 7.6). Splenocytes were washed twice with 10 mM Tris-HCl, pH 8.5, and the pellet was immediately frozen at −80 °C. Cell extracts were obtained by diluting the splenocyte pellet 1:1 with 10 mM Tris-HCl, pH 8.5, followed by mechanical lysis in a Potter–Elvehejm homogenizator (20 strokes, 1,000 r.p.m.). The lysate was centrifuged for 1 h at 200,000 × *g* at 4 °C and the supernatant was promptly used for further experiments.

Purification of active fractions from porcine splenocyte extract. Details of purification are given in Supplementary Fig. 1.

Mass spectrometry. Lyophilized active fractions were dissolved in 50% acetonitrile/water (liquid-chromatography–MS grade, Merck). Samples (1–2 μl) were applied to a gold-coated glass capillary (Type C nanoflow needle, Waters) and introduced into the nano-ESI source using a capillary voltage of 500–800 V, resulting in a spray of about 50 nl min^{-1}.

Mass detection was performed on a hybrid orthogonal acceleration QTOF MS/MS (Q-TOF Premier, Waters) equipped with an ESI interface and operated in a negative ionization mode. The mass spectrometer was operated in V-mode and tuned to a resolution of 10.000 full-width at half-maximum (FWHM) for singly charged ions. Prominent peaks in the survey mass spectrometry spectrum were manually selected for fragmentation in MS/MS mode using collision energies between 15 and 30 eV.

Affinity purification of TcdB. TcdB-specific monoclonal antibody 2CV (7 mg)[37] was coupled to a HiTrap NHS Sepharose column (GE Healthcare), according to the manufacturer's protocol. Approximately 4 mg of conventionally purified TcdB were applied to this column and washed with a buffer containing 50 mM Tris/HCl, pH 7.0 and 125 mM NaCl. TcdB was eluted with 0.1 M triethanolamin-HCl, pH 11, and immediately neutralized with a one-tenth volume of 1 M Tris-HCl, pH 7.5.

EPNP modification of TcdB. TcdB was transferred to a buffer containing 250 mM sodium acetate, 500 mM NaCl, 1 mM EDTA, 500 μM DTT and 5% DMSO (pH 5.5). EPNP (10 mM; Acros Organics, Belgium) was added and incubated for 30 h at 500 r.p.m. on an orbital mixer at room temperature. Finally the buffer was changed (ZebaSpin desalting colums, Pierce) to HEPES 40 mM, pH 7.5.

Microinjection experiments. Microinjection experiments were performed with IHKE cells maintained in DMEM/HAM's F12-medium (1:1) supplemented with 10% FCS and 2 mM glutamine in the presence of 65 mg l^{-1} penicillin-G and 100 mg l^{-1} streptomycin. IHKE cells were seeded overnight on 35 mm Petri dishes with glass bottoms and then used in the experiments. Native or EPNP-modified TcdB were diluted in an intracellular buffer (145 mM potassium glutamate, 20 mM NaCl, pH 7.3) to a final concentration of 1 μg ml^{-1}. The intracellular buffer contained fluorescein isothiocyanate (FITC)-labelled dextran (100 mg ml^{-1}, molecular mass approximately 19,500 Da; Sigma) to visualize successful microinjection. Microinjection was performed with an Eppendorf micromanipulator 5170 and microinjector 5242 system (Eppendorf AG). Injection pulses of 120 hPa were applied for 0.3 s. Images were taken at several time points (2 h, 7 h and 24 h) and acquired with an inverted Leitz DM-IRB phase-contrast microscope (Leica Microsystems) equipped with a Visitron 'VisiChrome' High Speed Fluorescence Polychromatorsystem (Visitron Systems). Excitation and emission wavelengths for fluorescence images were 490 and 525 nm.

For data acquisition, storage and processing we used Digital Camera Photometrics 'Coolsnap cf' (Roper Scientific) and Metafluor acquisition software (Universal Imaging Corporation).

Mass spectrometry of native and EPNP-modified TcdB. Details of the mass spectrometry are given in Supplementary Fig. 4.

Received 13 December 2006; accepted 25 January 2007.
Published online 4 March 2007.

1. Tenover, F. C. Mechanisms of antimicrobial resistance in bacteria. *Am. J. Infect. Control* **34,** S3–S10 (2006).
2. Fishman, N. Antimicrobial stewardship. *Am. J. Infect. Control* **34,** S55–S63 (2006).
3. Bartlett, J. G., Chang, T. W., Gurwith, M., Gorbach, S. L. & Onderdonk, A. B. Antibiotic-associated pseudomembranous colitis due to toxin-producing *Clostridia*. *N. Engl. J. Med.* **298,** 531–534 (1978).
4. George, R. H. *et al.* Identification of *Clostridium difficile* as a cause of pseudomembranous colitis. *Br. Med. J.* **1,** 695 (1978).
5. Gorbach, S. L. Antibiotics and *Clostridium difficile*. *N. Engl. J. Med.* **341,** 1690–1691 (1999).
6. Davey, P. *et al.* Systematic review of antimicrobial drug prescribing in hospitals. *Emerg. Infect. Dis.* **12,** 211–216 (2006).
7. Lyerly, D. M., Krivan, H. C. & Wilkins, T. D. *Clostridium difficile*: its disease and toxins. *Clin. Microbiol. Rev.* **1,** 1–18 (1988).
8. Voth, D. E. & Ballard, J. D. *Clostridium difficile* toxins: mechanism of action and role in disease. *Clin. Microbiol. Rev.* **18,** 247–263 (2005).
9. Eichel-Streiber, C., Laufenberg-Feldmann, R., Sartingen, S., Schulze, J. & Sauerborn, M. Comparative sequence analysis of the *Clostridium difficile* toxins A and B. *Mol. Gen. Genet.* **233,** 260–268 (1992).
10. Green, G. A., Schue, V. & Monteil, H. Cloning and characterization of the cytotoxin L-encoding gene of *Clostridium sordellii*: homology with *Clostridium difficile* cytotoxin B. *Gene* **161,** 57–61 (1995).
11. Hofmann, F., Herrmann, A., Habermann, E. & Eichel-Streiber, C. Sequencing and analysis of the gene encoding the α-toxin of *Clostridium novyi* proves its homology to toxins A and B of *Clostridium difficile*. *Mol. Gen. Genet.* **247,** 670–679 (1995).
12. Eichel-Streiber, C., Boquet, P., Sauerborn, M. & Thelestam, M. Large clostridial cytotoxins—a family of glycosyltransferases modifying small GTP-binding proteins. *Trends Microbiol.* **4,** 375–382 (1996).
13. Rupnik, M. *et al.* Revised nomenclature of *Clostridium difficile* toxins and associated genes. *J. Med. Microbiol.* **54,** 113–117 (2005).
14. Loo, V. G. *et al.* A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. *N. Engl. J. Med.* **353,** 2442–2449 (2005).
15. McDonald, L. C. *et al.* An epidemic, toxin gene-variant strain of *Clostridium difficile*. *N. Engl. J. Med.* **353,** 2433–2441 (2005).
16. Centers for Disease Control and Prevention (CDC). Severe *Clostridium difficile*-associated disease in populations previously at low risk—four states, 2005. *Morb. Mortal. Wkly Rep.* **54,** 1201–1205 (2005).
17. Kuijper, E. J., Coignard, B. & Tull, P. Emergence of *Clostridium difficile*-associated disease in North America and Europe. *Clin. Microbiol. Infect.* **12** (Suppl. 6), 2–18 (2006).
18. McDonald, L. C. *Clostridium difficile*: responding to a new threat from an old enemy. *Infect. Control Hosp. Epidemiol.* **26,** 672–675 (2005).
19. Eichel-Streiber, C., Sauerborn, M. & Kuramitsu, H. K. Evidence for a modular structure of the homologous repetitive C-terminal carbohydrate-binding sites of *Clostridium difficile* toxins and *Streptococcus mutans* glucosyltransferases. *J. Bacteriol.* **174,** 6707–6710 (1992).
20. Pfeifer, G. *et al.* Cellular uptake of *Clostridium difficile* toxin B. Translocation of the N-terminal catalytic domain into the cytosol of eukaryotic cells. *J. Biol. Chem.* **278,** 44535–44541 (2003).
21. Hofmann, F., Busch, C., Prepens, U., Just, I. & Aktories, K. Localization of the glucosyltransferase activity of *Clostridium difficile* toxin B to the N-terminal part of the holotoxin. *J. Biol. Chem.* **272,** 11074–11078 (1997).
22. Tucker, K. D. & Wilkins, T. D. Toxin A of *Clostridium difficile* binds to the human carbohydrate antigens I, X, and Y. *Infect. Immun.* **59,** 73–78 (1991).
23. Karlsson, K. A. Microbial recognition of target-cell glycoconjugates. *Curr. Opin. Struct. Biol.* **5,** 622–635 (1995).
24. Florin, I. & Thelestam, M. Internalization of *Clostridium difficile* cytotoxin into cultured human lung fibroblasts. *Biochim. Biophys. Acta* **763,** 383–392 (1983).
25. Florin, I. & Thelestam, M. Lysosomal involvement in cellular intoxication with *Clostridium difficile* toxin B. *Microb. Pathog.* **1,** 373–385 (1986).
26. Henriques, B., Florin, I. & Thelestam, M. Cellular internalisation of *Clostridium difficile* toxin A. *Microb. Pathog.* **2,** 455–463 (1987).
27. Giesemann, T. *et al.* Cholesterol-dependent pore formation of *Clostridium difficile* toxin A. *J. Biol. Chem.* **281,** 10808–10815 (2006).
28. Qa'Dan, M., Spyres, L. M. & Ballard, J. D. pH-induced conformational changes in *Clostridium difficile* toxin B. *Infect. Immun.* **68,** 2470–2474 (2000).
29. Rupnik, M. *et al.* Characterization of the cleavage site and function of resulting cleavage fragments after limited proteolysis of *Clostridium difficile* toxin B (TcdB) by host cells. *Microbiology* **151,** 199–208 (2005).
30. Sehr, P. *et al.* Glucosylation and ADP ribosylation of Rho proteins: effects on nucleotide binding, GTPase activity, and effector coupling. *Biochemistry* **37,** 5296–5304 (1998).
31. Just, I. *et al.* Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature* **375,** 500–503 (1995).
32. Moos, M. & Eichel-Streiber, C. Purification and evaluation of large clostridial cytotoxins that inhibit small GTPases of Rho and Ras subfamilies. *Methods Enzymol.* **325,** 114–125 (2000).
33. Hsu, F. F., Turk, J. & Gross, M. L. Structural distinction among inositol phosphate isomers using high-energy and low-energy collisional-activated dissociation tandem mass spectrometry with electrospray ionization. *J. Mass Spectrom.* **38,** 447–457 (2003).
34. Fernandez-Patron, C., Hardy, E., Sosa, A., Seoane, J. & Castellanos, L. Double staining of Coomassie blue-stained polyacrylamide gels by imidazole-sodium dodecyl sulfate-zinc reverse staining: sensitive detection of Coomassie blue-undetected proteins. *Anal. Biochem.* **224,** 263–269 (1995).
35. Shears, S. B. Assessing the omnipotence of inositol hexakisphosphate. *Cell. Signal.* **13,** 151–158 (2001).
36. Saiardi, A., Bhandari, R., Resnick, A. C., Snowman, A. M. & Snyder, S. H. Phosphorylation of proteins by inositol pyrophosphates. *Science* **306,** 2101–2105 (2004).
37. Sauerborn, M., Hegenbarth, S., Laufenberg-Feldmann, R., Leukel, P. & von Eichel-Streiber, C. Monoclonal antibodies discriminating between *Clostridium difficile* toxins A and B. *Int. J. Med. Microbiol.* **24** (Suppl.), 510–511 (1994).
38. Sebaihia, M. *et al.* The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nature Genet.* **38,** 779–786 (2006).
39. Rao, M. B., Tanksale, A. M., Ghatge, M. S. & Deshpande, V. V. Molecular and biotechnological aspects of microbial proteases. *Microbiol. Mol. Biol. Rev.* **62,** 597–635 (1998).
40. Just, I. & Gerhard, R. Large clostridial cytotoxins. *Rev. Physiol. Biochem. Pharmacol.* **152,** 23–47 (2004).
41. Barth, H. *et al.* Low pH-induced formation of ion channels by *Clostridium difficile* toxin B in target cells. *J. Biol. Chem.* **276,** 10670–10676 (2001).
42. Macbeth, M. R. *et al.* Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* **309,** 1534–1539 (2005).
43. Byrum, J., Jordan, S., Safrany, S. T. & Rodgers, W. Visualization of inositol phosphate-dependent mobility of Ku: depletion of the DNA-PK cofactor $InsP_6$ inhibits Ku mobility. *Nucleic Acids Res.* **32,** 2776–2784 (2004).
44. York, J. D., Guo, S., Odom, A. R., Spiegelberg, B. D. & Stolz, L. E. An expanded view of inositol signaling. *Adv. Enzyme Regul.* **41,** 57–71 (2001).
45. Braun, V. & von Eichel-Streiber, C. Virulence-associated mobile elements in *Bacilli* and *Clostridia*. In *Pathogenicity Islands and Other Mobile Elements* (American Society of Microbiology, Washington DC, 1999).
46. Braun, V., Hundsberger, T., Leukel, P., Sauerborn, M. & Eichel-Streiber, C. Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. *Gene* **181,** 29–38 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Müller, J. Forsch, A. Malotta, N. Robertz and E. Scholz for technical assistance; M. Moos, C. Srokowski and C. Tertilt for critically discussing the manuscript; and A. Lange for his advice in affinity purification of TcdB. Special thanks to M. Popoff for providing *C. sordellii* TcsL of high quality. This work was supported by grants from the Stiftung Innovation Rheinland-Pfalz and Deutsche Forschungs Gemeinschaft (to C.v.E.-S.). S.T. and H.S. were supported by the Sonderforschungsbereich 'Invasion und Persistenz bei Infektionen', the Hochschulbauförderungsgesetz Program (H.S.), and the Immunology Cluster of Excellence (ICE) at the University of Mainz (H.S.). M.R. was supported by an EMBO grant. C.v.E.-S. acknowledges the support of the University of Mainz for offering additional laboratory space in the Verfügungsgebäude für Forschung und Entwicklung.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.v.E.-S. (veichel@uni-mainz.de) or H.S. (schild@uni-mainz.de).

LETTERS

Acceleration of the rotation of asteroid 1862 Apollo by radiation torques

Mikko Kaasalainen[1], Josef Ďurech[2], Brian D. Warner[3], Yurij N. Krugly[4] & Ninel M. Gaftonyuk[5]

The anisotropic reflection and thermal re-emission of sunlight from an asteroid's surface acts as a propulsion engine. The net propulsion force (Yarkovsky effect) changes the orbital dynamics of the body at a rate that depends on its physical properties; for irregularly shaped bodies, the propulsion causes a net torque (the Yarkovsky–O'Keefe–Radzievskii–Paddack or YORP effect) that can change the object's rotation period and the direction of its rotation axis[1,2]. The Yarkovsky effect has been observed directly[3], and there is also indirect evidence of its role in the orbital evolution of asteroids over long time intervals[4–6]. So far, however, only indirect evidence exists for the YORP effect through the clustering of the directions of rotation axes in asteroid families[6–8]. Here we report a change in the rotation rate of the asteroid 1862 Apollo, which is best explained by the YORP mechanism. The change is fairly large and clearly visible in photometric lightcurves, amounting to one extra rotation cycle in just 40 years even though Apollo's size is well over one kilometre. This confirms the prediction that the YORP effect plays a significant part in the dynamical evolution of asteroids.

Although the YORP-induced change in axis direction is unlikely to be measurable in real time with ground-based astronomical techniques, the change in rotation rate is potentially detectable if the asteroid's shape, size and rotation direction are suitable[9]. One of the largest of such candidates is the asteroid 1862 Apollo, with an estimated average diameter of 1,400 m (refs 10, 11) (Supplementary Fig. S2). Apollo has a small satellite[12]: its size is less than one-tenth of the primary, so its influence on the asteroid's main dynamics is negligible in the present context and timeline. Using lightcurve inversion methods[13–15], we analysed an extensive data set of photometric observations of Apollo from apparitions in 1980, 1982, 1998 and 2005, consisting of 39 good-quality lightcurves (1,642 data points). Such coverage is sufficient to produce a unique and well-representative solution for the global shape and the rotation properties[16–18]. The 1980 and 1982 observations are collected in the Uppsala Asteroid Photometric Catalogue[19]; many of these were presented and analysed in ref. 20, indicating the same basic period we present here, though with some ambiguity. The longer timeline and wider coverage enabled us to obtain an unambiguous result. Our new data from 1998 and 2005 are summarized in Figs 1 and 2, Supplementary Table S1 and Supplementary Fig. S1.

We found that Apollo's spin behaviour cannot be explained with a simple assumption of a constant sidereal rotation period P, as Fig. 1 shows. To achieve a data fit down to the noise level, we had to extend the standard lightcurve inversion model by a linear increase in time t in the rotation speed $\omega = 2\pi/P$:

$$\omega(t) = \omega(t_0) + \upsilon(t - t_0) \qquad (1)$$

where $\upsilon = \mathrm{d}\omega/\mathrm{d}t$ (the upsilon stands for YORP) is a constant with dimensions of rad day^{-2}. The first-order equation (1) corresponds to the expected YORP behaviour, that is, essentially a linear secular trend in time, with only small fluctuations owing to orbital dynamics around the Sun[7] (see also Supplementary Fig. S3). In our extended formulation, the parameter υ was determined in χ^2-minimization

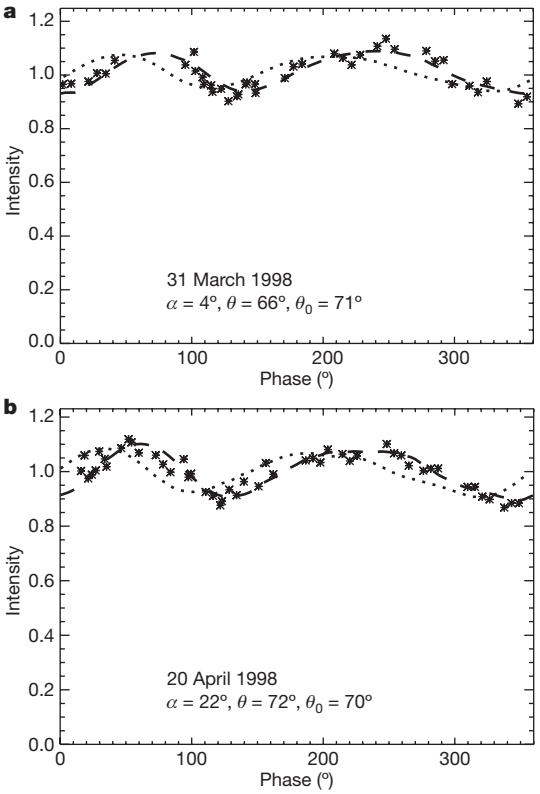


Figure 1 | YORP-induced phase offset in lightcurves. Lightcurves from the 1998 apparition are shown with asterisks, together with lightcurve inversion fits based on accelerated rotation (dashed line) and constant period (dotted line). The constant-period model phase offset in 1998 is very clear, particularly since the inclusion of υ fits the 1998 lightcurves to noise level. The viewing and illumination angles from the pole are given, respectively, by θ and θ_0, and the solar phase (Earth–Apollo–Sun) angle by α. The epochs of data points are folded to correspond to the rotational phase. Here the mean intensities are normalized to unity. The scattering model used was the Lommel-Seeliger/Lambert combination that produces results similar to the Hapke model[13]. The shape of 1862 Apollo from our full inverse problem solution is shown in Supplementary Fig. S2. As is typical for this type of an inverse problem, non-convex shape solutions did not yield better χ^2-values than a convex one. The lightcurve data do not contain proper information on non-convex details, so we adopt the convex shape version as the most stable and reliable one[15,16].

[1]Department of Mathematics and Statistics, Rolf Nevanlinna Institute, PO Box 68, FI-00014 University of Helsinki, Finland. [2]Astronomical Institute, Charles University, V Holešovičkách 2, 18000 Prague, Czech Republic. [3]Palmer Divide Observatory, Colorado Springs, Colorado 80908, USA. [4]Institute of Astronomy of Kharkiv National University, Sumska str. 35, Kharkiv 61022, Ukraine. [5]Crimean Astrophysical Observatory, Simeiz 98680, Ukraine.

simultaneously with all other parameters for spin—$\omega(t_0)$ and the ecliptic pole direction (β,λ)—and shape, by including it in ω in the rotation matrix that describes the orientation of the asteroid[13,15].

Although $\upsilon = 0$ (constant period) gives a clearly unsatisfactory fit to the data, the value $\upsilon = (5.3 \pm 1.3) \times 10^{-8}$ rad day^{-2} fits all data very well (Figs 1 and 2 and Supplementary Fig. S1), yielding a unique solution with the pole ($\beta = -71^\circ$, $\lambda = 50^\circ$) $\pm 7^\circ$ of arc and period $P(t_0) = 3.065447 \pm 0.000003$ h for Julian day epoch $t_0 = 2444557.0$ (13.5 November 1980) (Table 1). A constant-period fit yields a χ^2-value that is 15% higher, which alone is statistically sufficient to favour the solution with non-zero υ. We note that YORP affects the fit levels of lightcurves at separate apparitions differently (unlike pole direction and shape parameters, which have a much more uniform effect over the whole data set in fitting). In our case the data set is heavily dominated by data points at the beginning and end of the observing timeline, and the fits for these points are inevitably the least affected by YORP. The most revealing fit discrepancies with the constant-period solution occur in only a minority of the lightcurves. Thus it is essential to concentrate on this part of the data set to obtain a well-founded estimate for υ from both χ^2-levels and visual inspection.

The reason for YORP's uneven fit influence is depicted in Fig. 3. The linear change of equation (1) in ω leads to a phase lag of $\delta\phi(t)$ that is quadratic in time for the phase from rotation fixed at $\omega(t_0)$:

$$\delta\phi(t) = \tfrac{1}{2}\upsilon(t - t_0)^2 \qquad (2)$$

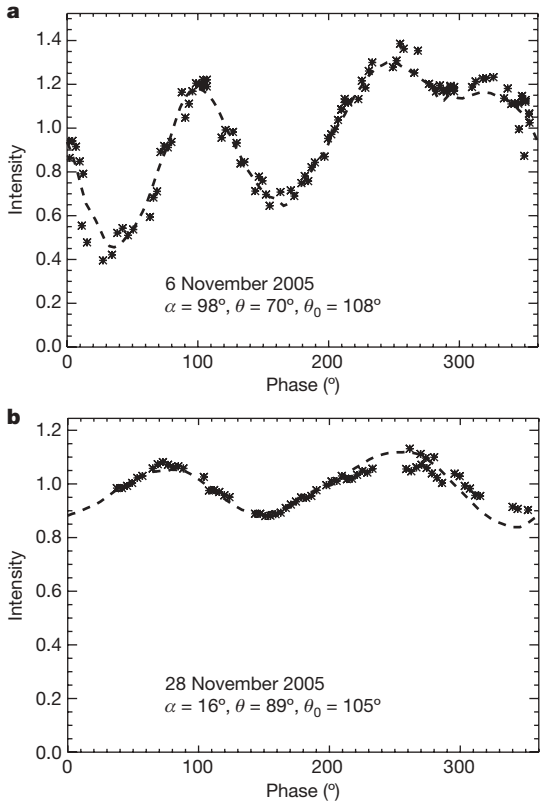


Figure 2 | Sample lightcurves from the 2005 apparition. Lightcurve points and inversion fits (dashed line) are shown as in Fig. 1. For clarity, we only show the full inversion fit as the constant-period fit is now very close to it (with slightly larger residuals). The brightness measurements from 1980 till 2005 cover Apollo's surface well from various viewing and illumination geometries, facilitating an accurate determination of Apollo's spin properties and global shape with lightcurve inversion methods in which the χ^2-value between the observed and modelled brightness points is minimized[13–15]. The robustness of these inverse problem techniques is based on mathematical uniqueness and stability theorems[16], and their validity has been well tested in practice with space probe missions[13], laboratory experiments[17] and other independent data[18].

Table 1 | Spin properties of 1862 Apollo

β	-71°
λ	50° (pole error $\pm 7^\circ$ of arc)
P (JD 2444557.0)	3.065447 ± 0.000003 h
υ	$(5.3 \pm 1.3) \times 10^{-8}$ rad day^{-2}
dP/dt	$(-1.4 \pm 0.3) \times 10^{-10} = (-1.2 \pm 0.3) \times 10^{-6}$ h year^{-1}

To estimate the error limits of the deduced υ value, we use the same method as for estimating the period error in ordinary lightcurve analysis, with both χ^2-level and phase comparison[13]. In period determination, local χ^2-minima in the trial period are spaced with the interval $\delta P = 0.5P^2/\Delta T$, which corresponds to a 180° phase offset at T. The period error is usually a small fraction of δP, typically amounting to an offset of about 5–8° at T because a larger error would cause both visible fit offsets as well as raise the χ^2-level significantly. Adopting the same principle, we estimate the error in υ to be 1.3×10^{-8} rad day^{-2}, corresponding to a phase uncertainty of 7° in 1998. In terms of the rate of change of the period $P(t)$, υ can be rewritten as the dimensionless quantity $\frac{dP}{dt} = -\frac{1}{2\pi}\upsilon P^2$.

This parabola for the nominal value of υ is shown as solid line in Fig. 3. On the other hand, any two local regions of any observing timeline can always be fitted with a constant-period model with negligible phase offsets. In our case, the data set is dominated by the two epoch regions whose corresponding locations on the parabola are connected by a dashed straight line. A constant-period fitting procedure tries to find an $\omega' > \omega(t_0)$ such that the phase offsets between the model and the data are minimized over the timeline, so the slope of the dashed line must be approximately $\omega' - \omega(t_0)$. The apparent model offset is now given by the distance between the dashed and solid lines, and in our case the crucial region of interest is 1998. There the apparent offset is 25°, in reality corresponding to the already quite sizeable but unseen $\delta\phi$ of 125° in 2005 (also note that the apparent phase offset in 1998 is of opposite sign to the actual $\delta\phi$). Apparent YORP offsets are thus both strongly attenuated and unevenly distributed. Because the offset is so obvious here, we can use both the phases from the constant-period fit and the χ^2-distribution from the full acceleration fit to ascertain the mutually consistent value and error estimate $\upsilon = (5.3 \pm 1.3) \times 10^{-8}$ rad day^{-2}.

To verify that our solution is consistent with the physics of the YORP mechanism, we used the obtained shape and spin model to estimate the strength of YORP on such a body[21,22]. Our initial

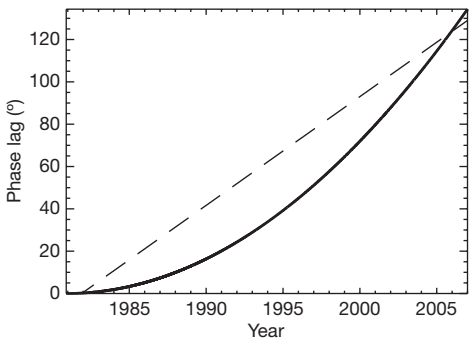


Figure 3 | Phase offset evolution in time. The phase evolution $\delta\phi$ for the observed value of υ, compared against phase from $\omega(t_0)$, is shown by the solid line, together with the corresponding evolution from the best-fitting constant-period solution $\omega' > \omega(t_0)$ (dashed line). The ratio of the apparent model offset (the difference between the solid and dashed lines) $\delta\phi_{app}(t)$ at time t to the actual full YORP offset $\delta\phi(T)$ at the end T of the observing timeline is: $\left|\frac{\delta\phi_{app}(t)}{\delta\phi(T)}\right| = \frac{\Delta t}{\Delta T}\left(1 - \frac{\Delta t}{\Delta T}\right)$ where $\Delta x = x - t_0$. This ratio is at most 1/4 when $\Delta t = \Delta T/2$; here it is 1/5 for 1998. The simulations of the expected period evolution of our inversion model (Supplementary Fig. S3) were consistent with the observations. For these simulations, the orbital ephemerides for Apollo were obtained by direct integration using OrbFit9 (http://newton.dm.unipi.it/orbfit/) software with necessary perturbations included. For the attitude dynamics we constructed a complete numerical model integrating the Euler equations. We included solar gravitational torque and YORP torques as in refs 21 and 22. Although the surface of Apollo has non-zero thermal inertia[10], we find that this parameter has little effect on YORP strength. In particular, we ran several simulations with the surface thermal conductivity K ranging from 0 to 0.1 W m^{-1} K^{-1} and found no difference in the results. This confirms previous findings[22].

simulation used a bulk density $\rho = 2.5\,\mathrm{g\,cm^{-3}}$ and scaled the nominal shape model to correspond to a body with equivalent size $D =$ 1,400 m. The predicted mean YORP acceleration of Apollo's rotation from our model was $(\mathrm{d}\omega/\mathrm{d}t)_{\mathrm{YORP}} = 4.6 \times 10^{-8}\,\mathrm{rad\,day^{-2}}$. Being aware of the high sensitivity of the computed YORP value on the details of the asteroid's shape, we also derived eight shape (and pole) variants very close to the nominal one that all fit the observed lightcurves within their uncertainty. We computed the YORP strength for each of these and observed a variation within $(-0.3$ to $+2.9) \times 10^{-8}\,\mathrm{rad\,day^{-2}}$ about the nominal value (Supplementary Fig. S3).

To make the central value $(\mathrm{d}\omega/\mathrm{d}t)_{\mathrm{YORP}}$ of the predicted YORP strength on Apollo coincide with the observed acceleration υ of its rotation rate, we note that $(\mathrm{d}\omega/\mathrm{d}t)_{\mathrm{YORP}} \propto (\rho D^2)^{-1}$. Hence $(\mathrm{d}\omega/\mathrm{d}t)_{\mathrm{YORP}} = \upsilon$ may be obtained for the nominal solution by using a bulk density of $2.2\,\mathrm{g\,cm^{-3}}$, rescaling the size to a slightly smaller value, or employing any other suitable density–size combination in between. Such values of effective bulk density for a Q-type asteroid such as Apollo are reasonable and lie within the interval of values found for other asteroids by other astronomical means[23]. Some of the variant solutions match the observation without any need of parameter readjustment. We thus conclude that the obtained model is self-consistent and the observed acceleration υ of Apollo's rotation rate can be completely explained with the YORP mechanism. We note that observing the YORP strength also provides a means of estimating the masses (densities) of small asteroids if their sizes and shapes are known sufficiently well[3,9].

Our result suggests a current characteristic timescale of $T_{\mathrm{YORP}} \approx \omega/(\mathrm{d}\omega/\mathrm{d}t) \approx 2.6$ Myr for YORP to significantly change the rotation rate in the future, that is, with the current orbit and spin state the rotation period would roughly decrease to one-half ($\sim$1.5 h) within T_{YORP}. This value is well beyond the break-up limit for cohesionless bodies, suggesting that YORP may cause structural changes, mass shedding or even fission of Apollo at some point in the future.

Using dynamical simulations, we determined a median dynamical lifetime of $T_{\mathrm{dyn}} \approx 10$ Myr for 1862 Apollo before its removal from the simulation by solar or planetary impacts (Supplementary Fig. S4). Apollo's obliquity of $\sim 160^\circ$ places the spin state outside the reach of important secular spin-orbit resonances[24,25]. We thus expect that future obliquity evolution will reflect only minor variations of the orbital inclination (the median orbital inclination at 5 Myr was only $\sim 10^\circ$ in our simulation), making YORP steadily accelerate Apollo's rotation rate at about the currently observed value. Because T_{dyn} is significantly longer than T_{YORP}, we have further corroboration for the suggestion above that YORP will play an important role in Apollo's future evolution by bringing it towards the critical rotation regime. Planetary close encounters are naturally a possible source of significant dynamical change, but their portion of the potential orbital fates is a minor one.

Given the dynamically evolved orbit of Apollo within the planet-crossing zone we speculate that this asteroid might already have approached the critical rotation limit by YORP in the past. One possible way to remove the excess rotational angular momentum is to shed mass, which could have produced Apollo's small satellite. Such binary formation via YORP is a mechanism predicted and simulated by several studies[26–28]. Other 'landslide'-type shape rearrangements (changing the YORP behaviour even to a deceleration of rotation) or complete escape of material are also possible because the size of the (remaining) satellite is too small to contain a significant fraction of angular momentum.

Received 5 December 2006; accepted 15 January 2007.
Published online 7 March 2007.

1. Rubincam, D. P. Radiative spin-up and spin-down of small asteroids. *Icarus* **148,** 2–11 (2000).
2. Bottke, W. F., Vokrouhlický, D., Rubincam, D. P. & Nesvorný, D. The Yarkovsky and YORP effects: Implications for asteroid dynamics. *Annu. Rev. Earth Planet. Sci.* **34,** 157–191 (2006).
3. Chesley, S. R. *et al.* Direct detection of the Yarkovsky effect via radar ranging to the near-Earth asteroid 6489 Golevka. *Science* **302,** 1739–1742 (2003).
4. Bottke, W. F. *et al.* Dynamical spreading of asteroid families via the Yarkovsky effect: The Koronis family and beyond. *Science* **294,** 1693–1696 (2001).
5. Morbidelli, A. & Vokrouhlický, D. The Yarkovsky-driven origin of near Earth asteroids. *Icarus* **163,** 120–134 (2003).
6. Vokrouhlický, D. *et al.* Yarkovsky/YORP chronology of asteroid families. *Icarus* **182,** 118–142 (2006).
7. Vokrouhlický, D., Nesvorný, D. & Bottke, W. F. The vector alignments of asteroid spins by thermal torques. *Nature* **425,** 147–152 (2003).
8. La Spina, A., Paolicchi, P., Kryszczyńska, A. & Pravec, P. Retrograde spins of near-Earth asteroids from the Yarkovsky effect. *Nature* **428,** 400–401 (2004).
9. Vokrouhlický, D., Čapek, D., Kaasalainen, M. & Ostro, S. J. Detectability of YORP rotational slowing of asteroid 25143 Itokawa. *Astron. Astrophys.* **414,** L21–L24 (2004).
10. Harris, A. W. A thermal model for near-Earth asteroids. *Icarus* **131,** 291–301 (1998).
11. Binzel, R. P., Lupishko, D. F., Di Martino, M., Whiteley, R. J. & Hahn, G. J. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 255–271 (Univ. Arizona Press, Tucson, 2003).
12. Ostro, S. J. *et al.* 1862 Apollo. *IAU Circ.* **8627** (2006).
13. Kaasalainen, M., Torppa, J. & Muinonen, K. Optimization methods for asteroid lightcurve inversion. II. The complete inverse problem. *Icarus* **153,** 37–51 (2001).
14. Kaasalainen, M., Mottola, S. & Fulchignoni, M. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 139–150 (Univ. Arizona Press, Tucson, 2003).
15. Kaasalainen, M. & Ďurech, J. in *Near Earth Objects, our Celestial Neighbors: Opportunity and Risk* (eds Milani, A. Valsecchi, G. B. & Vokrouhlický, D.) (Cambridge University Press, Cambridge, in the press).
16. Kaasalainen, M. & Lamberg, L. Inverse problems of generalized projection operators. *Inverse Problems* **22,** 749–769 (2006).
17. Kaasalainen, S., Kaasalainen, M. & Piironen, J. Ground reference for space remote sensing: Laboratory photometry of an asteroid model. *Astron. Astrophys.* **440,** 1177–1182 (2005).
18. Marchis, F. *et al.* Shape, size and multiplicity of main-belt asteroids: I. Keck adaptive optics. *Icarus* **185,** 39–63 (2006).
19. Lagerkvist, C.-I., Piironen, J. & Erikson, A. *Asteroid Photometric Catalogue, 5th Update* (Uppsala Univ. Press, Uppsala, 2001).
20. Harris, A. W. *et al.* Photoelectric lightcurves of the asteroid 1862 Apollo. *Icarus* **70,** 246–256 (1987).
21. Vokrouhlický, D. & Čapek, D. YORP-induced long-term evolution of the spin state of small asteroids and meteoroids. I. Rubincam's approximation. *Icarus* **159,** 449–467 (2002).
22. Čapek, D. & Vokrouhlický, D. The YORP effect with finite thermal conductivity. *Icarus* **172,** 526–536 (2004).
23. Britt, D. T., Yeomans, D., Housen, K. & Consolmagno, G. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 485–500 (Univ. Arizona Press, Tucson, 2003).
24. Vokrouhlický, D., Bottke, W. F. & Nesvorný, D. The spin state of 433 Eros and its possible implications. *Icarus* **175,** 419–434 (2005).
25. Vokrouhlický, D., Nesvorný, D. & Bottke, W. F. Secular spin dynamics of inner main-belt asteroids. *Icarus* **184,** 1–28 (2006).
26. Paddack, S. J. Rotational bursting of small celestial bodies: Effects of radiation pressure. *J. Geophys. Res.* **74,** 4379–4381 (1969).
27. Bottke, W. F., Vokrouhlický, D., Rubincam, D. P. & Brož, M. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 395–408 (Univ. Arizona Press, Tucson, 2003).
28. Ostro, S. J. *et al.* Radar imaging of binary near-Earth asteroid (66391) 1999 KW4. *Science* **314,** 1276–1280 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Vokrouhlický for assistance and discussions, P. Pravec for Apollo data, and W. Bottke and A. Harris for comments. The work of M.K. was supported by the Academy of Finland, that of J.Ď. by the Grant Agency of the Czech Republic, and that of Y.N.K. and N.M.G. by the Ministry of Education and Science of the Ukraine.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.K. (mikko.kaasalainen@helsinki.fi).

LETTERS

Biasing reaction pathways with mechanical force

Charles R. Hickenboth[1], Jeffrey S. Moore[1,2,3], Scott R. White[3,4], Nancy R. Sottos[2,3], Jerome Baudry[1] & Scott R. Wilson[1]

During the course of chemical reactions, reactant molecules need to surmount an energy barrier to allow their transformation into products. The energy needed for this process is usually provided by heat, light, pressure or electrical potential, which act either by changing the distribution of the reactants on their ground-state potential energy surface or by moving them onto an excited-state potential energy surface and thereby facilitate movement over the energy barrier. A fundamentally different way of initiating or accelerating a reaction is the use of force to deform reacting molecules along a specific direction of the reaction coordinate. Mechanical force has indeed been shown to activate covalent bonds in polymers, but the usual result is chain scission[1]. Here we show that mechanically sensitive chemical groups make it possible to harness the mechanical forces generated when exposing polymer solutions to ultrasound[2], and that this allows us to accelerate rearrangement reactions and bias reaction pathways to yield products not obtainable from purely thermal or light-induced reactions. We find that when placed within long polymer strands, the *trans* and *cis* isomers of a 1,2-disubstituted benzocyclobutene undergo an ultrasound-induced electrocyclic ring opening in a formally conrotatory and formally disrotatory process, respectively, that yield identical products. This contrasts with reaction initiation by light or heat alone[3], in which case the isomers follow mutually exclusive pathways to different products. Mechanical forces associated with ultrasound can thus clearly alter the shape of potential energy surfaces[4] so that otherwise forbidden or slow processes proceed under mild conditions, with the directionally specific nature of mechanical forces providing a reaction control that is fundamentally different from that achieved by adjusting chemical or physical parameters. Because rearrangement in our system occurs before chain scission, the effect we describe might allow the development of materials that are activated by mechanical stress fields.

We used the COGEF (COnstrained Geometries simulate External Force) method[5] computationally to determine the effects of molecular deformation on *trans* and *cis* 1,2-dimethoxybenzocyclobutenes. A profile of energy change relative to the ground state versus molecular elongation was produced for each isomer by systematically increasing the distance *d* between the methyl carbons while allowing all other atoms to fully relax (Fig. 1a). The energy and geometry at each deformation state was calculated at the B3LYP/6-31G* level of density functional theory. The calculations predict that the two isomers are distorted in a different rotational sense. In the *trans* isomer, the simulated extensional deformation induces a conrotatory torquing motion of the substituents, consistent with the atomic motions experienced for the thermally allowed pathway for ring opening[3] (Fig. 1b). In contrast, strain induces a disrotatory torquing motion of the substituents in the *cis* isomer[6], consistent with the Woodward–Hoffman allowed photochemical pathway for ring opening. These simulations predict that the mechanically activated processes are proceeding along thermally allowed and disallowed pathways for the *trans* and *cis* isomers, respectively[3]. Assuming these calculations accurately describe the geometric and electronic changes that accompany the action of extensional deformation, a formal disrotatory ring opening is predicted for *cis* benzocyclobutene (BCB). The presence of reaction products that result from the conrotatory ring opening of the *trans* isomer and from disrotatory ring opening of the *cis* isomer would strongly indicate that a mechanochemical process is in operation.

To test these predictions experimentally, we incorporated bifunctional BCB units **1** and **2** into well-defined, link-functionalized polymers[7] (LFPs) (Fig. 2). Using this preparation, only a single mechanophore—the putative mechanochemically active unit—is incorporated into the centre of each polymer chain. The incorporation of the mechanophores into polymers is necessary to study their mechanochemical behaviour because small molecules are largely inert to mechanical forces[8]. To carry out the synthesis, the carboxylic acid functions of **1** or **2** were coupled to α-methoxy-ω-amino poly (ethylene glycol) (mPEG)[9]. This synthesis yielded a series of polymers **3** and **4**, which contained the *trans*-BCB linker and the *cis*-BCB linker respectively, in weight-averaged molecular masses of M_w = 4, 10, 20, 40 and 60 kDa.

Having only a single BCB unit per 40 kDa polymer chain requires sensitive analytical techniques to observe reactivity at this specific site. *N*-(1-pyrene)-maleimide **5** (ref. 10) was used as a dienophile trap; the pyrene label allows the reaction to be monitored by gel permeation chromatography (GPC) using an ultraviolet (UV) detector. The strength of the UV signal is normalized against that obtained from a refractive index (RI) detector. Each polymer molecule contains only a single BCB unit, so cycloaddition of dienophile **5** and the *ortho*-quinodimethide (*o*QDM) diene generated by ring opening will result in one pyrene chromophore per polymer chain (Fig. 2).

It has previously been demonstrated that the application of ultrasound to dilute solutions of LFPs can be used to fragment a polymer chain at a single specific site[9]. The action of ultrasound on polymer molecules is generally thought to be mechanical, owing to frictional forces generated by the relative movement of solvent and macromolecules resulting from collapse of cavitation bubbles[2]. A dilute solution of the 40 kDa LFP **3** or **4** was sonicated in acetonitrile in the presence of a 500-fold excess of **5**. After sonication, each polymer was analysed by GPC, monitoring the UV absorbance at 345 nm, a wavelength at which the starting polymer has no absorbance but the pyrene label strongly absorbs. It was found that the UV signal significantly increased relative to the RI signal in both polymers **3** and **4** (Fig. 3a and b). This result suggests that the pyrene-labelled dienophile was incorporated into both LFPs, a particularly interesting result in the case of **4** given the calculated predictions. Significantly, these results were obtained for reactions conducted at approximately 6–9 °C, a temperature range for which the rate of thermal ring opening is negligible.

Several control experiments were carried out. There was no increase in UV absorbance when LFP **3** was manipulated exactly as described

[1]School of Chemical Sciences, [2]Department of Materials Science and Engineering, [3]The Beckman Institute, [4]Department of Aerospace Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

above, with the ultrasound omitted (Supplementary Information). This control experiment indicates that ultrasound is necessary for the observed response. A PEG homopolymer of M_w 43 kDa containing no BCB linker was subjected to the same conditions as above. Although the UV absorbance increased slightly, it was by far less than for **3** or **4** (Fig. 3c). These observations indicate that the increased UV absorbance was due to a mechanically activated ring opening reaction, followed by cycloaddition of the *o*QDM intermediate with **5**.

Previous studies on ultrasound-induced polymer degradation have demonstrated a chain-length dependence on the rate of reaction[2,8,9].

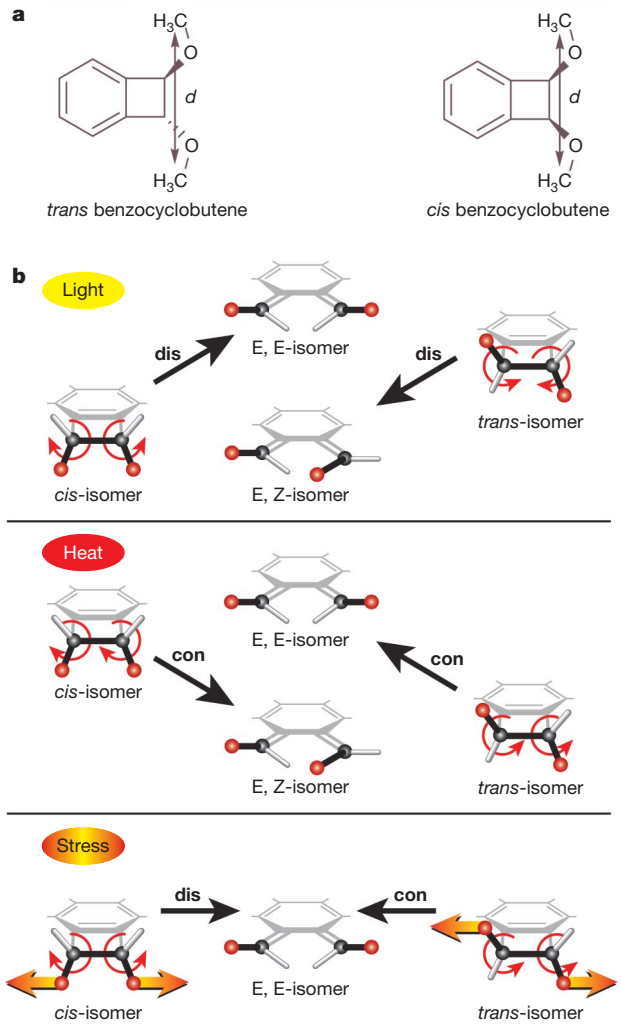


Figure 1 | Electrocyclic ring opening of benzocyclobutenes. The intermediate resulting from the electrocyclic ring opening of benzocyclobutenes depends on the method of activation and on the geometry of the molecule. **a**, To simulate mechanical tensile force, the distance *d* between the terminal methyl carbons of *trans* and *cis* 1,2-dimethoxybenzocyclobutene was systematically increased. At each step the geometry was minimized and the energy was calculated at the density functional theory B3LYP/6-31G* level. **b**, The expected pathways of photochemically activated and thermally activated electrocyclic ring opening of BCB, and the computationally predicted results of mechanically activated ring opening. Upon activation by light, the Woodward–Hoffman rules predict that both *cis* and *trans* BCB undergo disrotatory ('dis') ring opening. The *cis* isomer yields the E,E-*o*QDM intermediate, and the *trans* isomer yields the E,Z-*o*QDM intermediate (with *zusammen* double bond geometry). Upon activation by heat, both isomers undergo conrotatory ('con') ring opening. The *cis* isomer yields the E,Z-*o*QDM intermediate, and the *trans* isomer yields the E,E-oQDM intermediate. Activation by mechanical forces is predicted to induce a formal disrotatory ring opening in the *cis* isomer and a formal conrotatory ring opening in the *trans* isomer, such that both isomers are predicted to yield the same E,E-*o*QDM intermediate.

In particular, longer chains experience larger solvodynamic shear forces and accelerated reaction rates. Assuming that force reduces the kinetic barrier to ring opening, larger forces will correlate to faster rates. A reaction promoted purely by thermal energy is not expected to have such a M_w dependence. The LFPs **3** and **4** with M_w = 4, 10, 20 and 60 kDa were sonicated in the presence of **5**. The UV/RI signal ratios at the elution time corresponding to the M_w of the LFP were monitored by GPC. It was found that for both LFPs **3** and **4**, the UV/RI signal ratio increased smoothly with increasing M_w above a threshold value of about 20 kDa (Fig. 3d). The increasing response with increasing M_w indicates that the reaction in both polymers is activated by a mechanical process. Interestingly, for a given M_w, *cis* LFP **4** was found to react to a greater extent than *trans* LFP **3** for all the M_w values studied above 20 kDa.

We also performed control experiments with PEG homopolymers of M_w = 5, 10, 20, 26 and 60 kDa. In this case, the UV/RI ratio of the unbroken chains[11] is independent of M_w (Fig. 3d). Furthermore, a 1:1 mol ratio mixture of 4 kDa LFP **4** and 43 kDa PEG was sonicated. No significant increase in the UV signal was observed (Supplementary Information). These results suggest that the linker must be embedded within a polymer and the M_w must be above the critical value to observe reactivity: mixing alone is not enough.

To determine the specific chemical structures of the pyrene-modified LFPs, and to investigate the stereochemical consequences of mechanical activation, we resynthesized pyrene-labelled maleimide **5**, enriched with ^{13}C isotope in the carbonyl positions. Polymers **3** and **4** (40 kDa) were subjected to ultrasound in the presence of **5**^{13}**C**. After sonication, each polymer was analysed by ^{13}C nuclear magnetic resonance (NMR). It was found that both polymers **3** and **4** yielded a single resonance at δ 174.2 p.p.m. (Fig. 4a). That the polymers provided the same product was confirmed by examination of mixtures. When 1:1 mixtures of sonicated LFPs **3** and **4** were measured in either $CDCl_3$ or CD_3OD, only a single ^{13}C carbonyl peak was observed. Comparison of the ^{13}C NMR spectra obtained after sonication of LFPs **3** and **4** to model compound **6** indicated that the observed NMR signal is due to the Diels–Alder adduct, which results from endo addition of the maleimide to the E,E-*o*QDM intermediate (with *entgegen* double bond geometry) (Fig. 4b). Sonication of 43 kDa PEG under the same conditions yielded no detectable resonances in the carbonyl region.

The action of ultrasound on the *trans* BCB in polymer **3** induces a formal conrotatory ring opening to the *o*QDM intermediate, while the *cis* BCB in polymer **4** reacts via a formal disrotatory ring opening. That different modes of reactivity are observed for the two different isomers confirms that the activation is neither thermal nor photochemical but rather behaves according to COGEF simulations of mechanical activation.

The consequence of ultrasound-induced reactivity of LFPs **3** and **4** was compared to the products obtained under purely thermal conditions. The thermal reaction products were generated by reacting trap **5**^{13}**C** with polymer **3** in toluene at 105 °C and with polymer **4** in xylenes at 140 °C (ref. 12). ^{13}C analysis indicated that *trans* LFP **3** reacted smoothly with maleimide **5**^{13}**C** to give the same major product observed with polymer **3** and **4** under ultrasound conditions (Fig. 4a). The thermal reaction of polymer **4** did not yield any recognizable products. This result is consistent with previous reports[12] and our own experience on small-molecule model compounds that *cis*-1,2-diacetoxybenzocyclobutenes thermally degrade to a complex mixture.

To determine whether mechanical activation of the BCB linker may occur without polymer chain scission, polymer **3** was sonicated in the presence of maleimide trap **5**^{13}**C**, yielding a bimodal GPC trace with peak molecular weights of 40 and 20 kDa, presumably corresponding to unbroken and broken chains. The 40 kDa polymer, separated from the 20 kDa byproduct by preparatory GPC, showed the characteristic UV increase of pyrene incorporation by cycloaddition. Furthermore, ^{13}C analysis revealed that the Diels–Alder adduct was

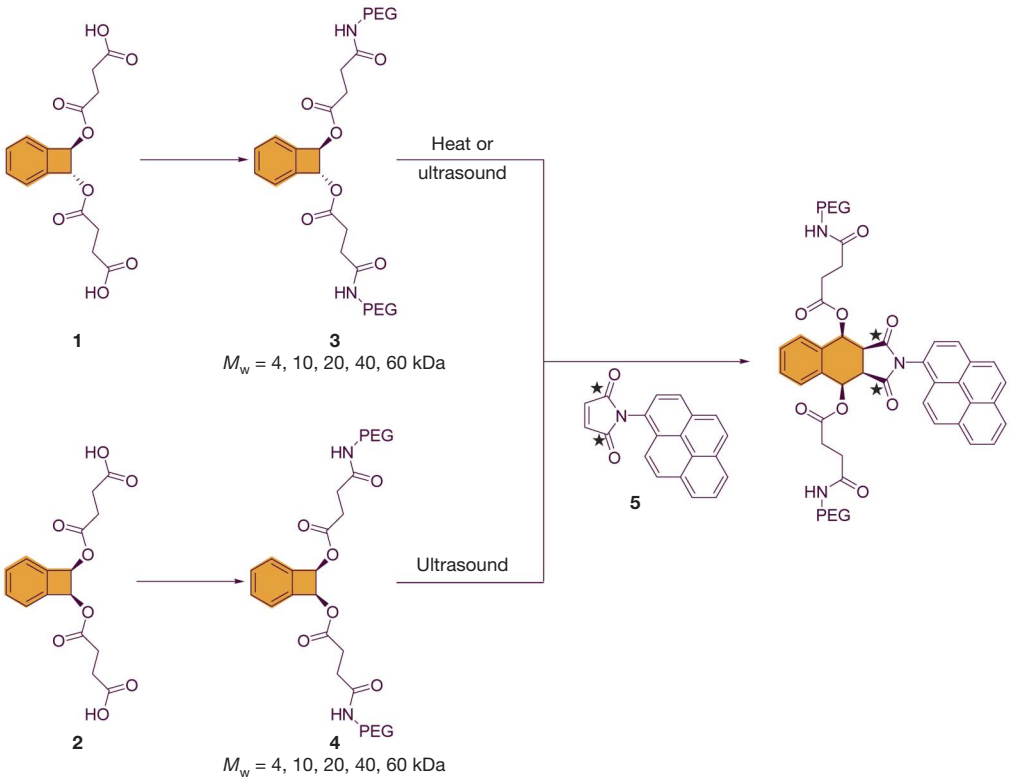


Figure 2 | Preparation and reaction of mechanosensitive polymers. Shown are the preparation of link-functionalized polymers **3** and **4** from bifunctional BCB units **1** and **2**, and the reaction of these polymers with maleimide trap **5**. For $\mathbf{5}^{\mathbf{13}}\mathbf{C}$, the positions of the ^{13}C labels are marked with a star.

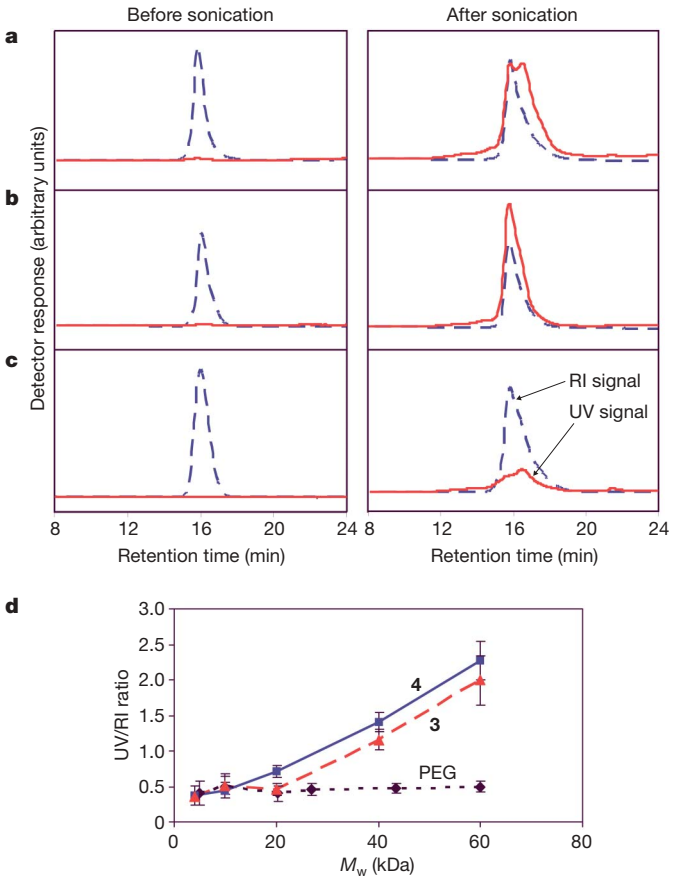


Figure 3 | Reaction of the mechanosensitive polymers. a–c, The top panels show GPC traces of the polymers before (left) and after (right) sonication. **a**, The 40 kDa *trans* polymer **3**. **b**, The 40 kDa *cis* polymer **4**. **c**, The 43 kDa PEG polymer. **d**, A plot of the ratio of the UV absorption to refractive index signal intensity versus polymer M_w for **3**, **4** and PEG. The increased UV absorption of LFPs **3** and **4** as a function of M_w is evidence that the ultrasound-induced reaction proceeds via mechanical activation. Error bars indicate the standard deviation observed for 2–4 independent trials.

present in the uncleaved polymer strands (Supplementary Information). This observation demonstrates that the LFPs can be mechanically activated without fragmenting the backbone.

The findings can be summarized by considering the reaction coordinate diagrams shown in Fig. 4c and d. Mechanical forces perform work[8] on the *trans* BCB unit by deforming bond lengths and angles, which reduces the barrier of a formal conrotatory process and accelerates the reaction (Fig. 4c). For the *cis* isomer, mechanical force does work on the BCB unit to reduce the barrier for a formal disrotatory process, even though it is known to be the higher energy ring opening pathway[3,6] (Fig. 4d). This explanation is consistent with the computational studies, which suggested that elongational deformation would induce a disrotatory motion for the *cis* isomer rather than the thermally allowed conrotatory ring opening (Supplementary Information). The disrotatory process is favoured because the molecular deformation induced by force is nearly congruent with the deformation along the disrotatory reaction pathway.

In conclusion, we have shown that ultrasound can be applied to polymer solutions to accelerate and alter the course of chemical reactions. The effects of ultrasound are consistent with the intuitive effects of a mechanical force, which acts to bias the reaction pathway towards products that best relieve the applied force by directly altering the molecular potential energy surface. Although there are practical limitations to the use of mechanochemistry for controlling the outcome of chemical reactions, we expect that the procedures

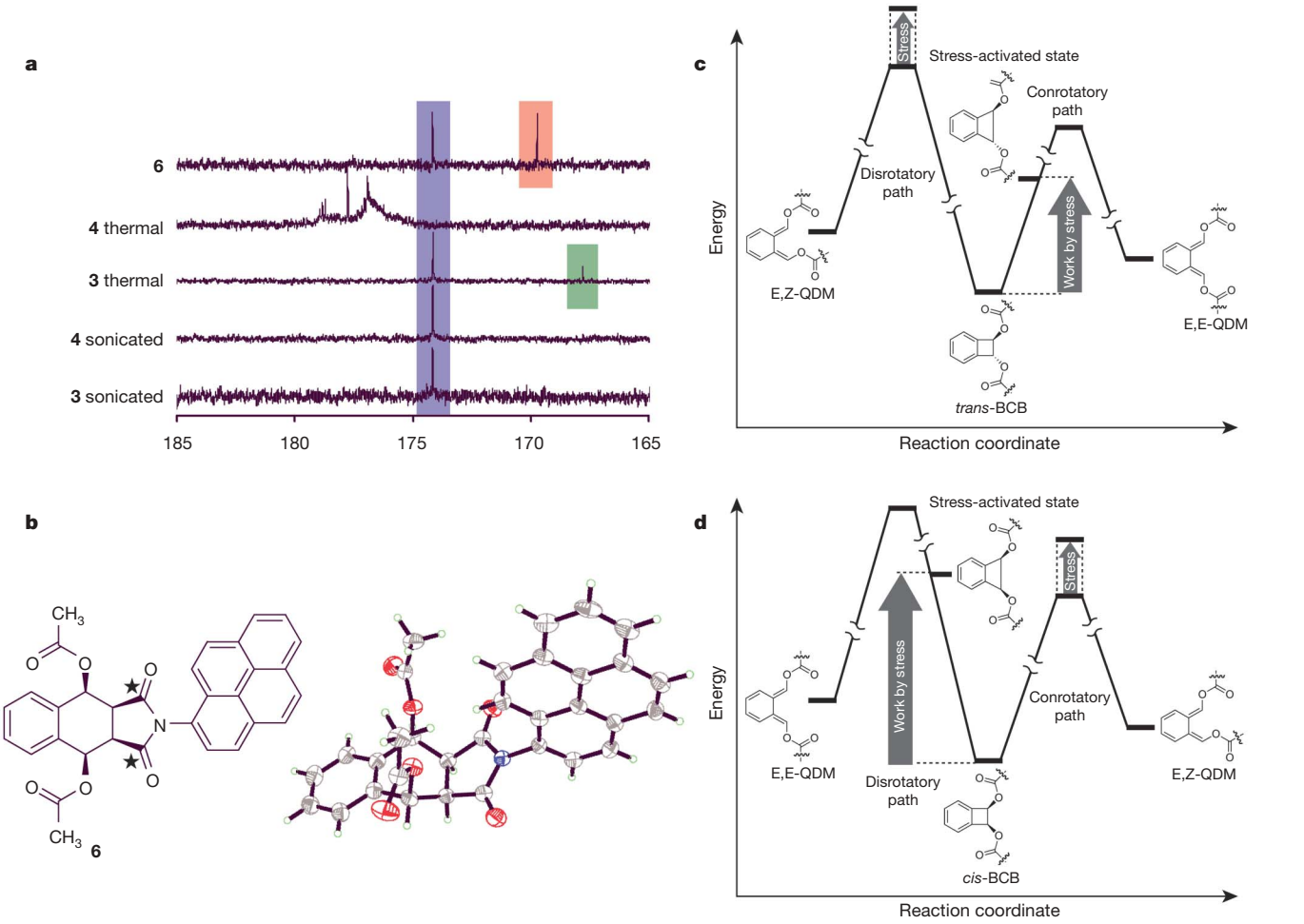


Figure 4 | Comparison of reaction products and pathways for *trans* and *cis* benzocyclobutenes induced by ultrasound and by heating. **a**, A portion of the ^{13}C NMR spectrum comparing the products formed from polymers **3** and **4** in the presence of **5**^{13}**C** by ultrasound and by thermal activation. Compound **6**, whose structure is confirmed by X-ray crystallography (**b**), exhibits a resonance at δ 174.2 p.p.m. (coloured blue) assigned to the Diels–Alder adduct resulting from endo addition of the maleimide dienophile to the E,E-*o*QDM intermediate (the carbon corresponding to the resonance at δ 174.2 is marked with a star). This resonance is not observed in the thermal reaction of polymer **4**. The resonance at δ 169.7 p.p.m. (coloured red) is assigned to the ester carbonyl in **6** at natural abundance. The resonance at δ 168.0 p.p.m. (coloured green) is due to *N*-pyrene-2,3-naphthimide, which occurs after thermal elimination of acid groups from the Diels–Alder adducts. **c**, In *trans*-BCB, mechanical force accelerates a formal conrotatory ring opening to the E,E-*o*QDM intermediate. **d**, In *cis*-BCB, mechanical force promotes a formal disrotatory ring opening to the E,E-*o*QDM intermediate even though the thermally allowed process is the conrotatory pathway. Additional transition states and intermediates may exist and are not shown[16].

reported here will be generally useful for discovering new mechanophores, the chemical reactivities of which are activated by external force.

METHODS

Sonication of polymers. To a solution of polymers **3** or **4** (0.75 mg ml^{-1}, 19 μM for 40 kDa polymer) in acetonitrile we added maleimide trap **5** (2.75 mg ml^{-1}). Approximately 12 ml of this solution was introduced into a Suslick cell[13]. A thermocouple and an argon line were threaded through the septa and placed in contact with the solution, ensuring that they did not touch the probe. The third side arm of the cell contained a septum. The cell was then wrapped in aluminium foil to exclude light. Argon was bubbled through the solution for 30 min before each experiment as well as during the experiment. The entire system was placed in an ice bath to maintain a temperature of 6–9 °C throughout sonication. The sample was then sonicated (pulsed on 0.5 s, off 1.0 s, at a frequency of 20 kHz and at an intensity of 8.73 W cm^{-2}) for a total time of 45 min (15 min sonication time). After sonication, the solvent was removed and the polymer sample was isolated by preparatory GPC. The concentrations used in all experiments were estimated to be below the critical overlap concentration for PEG[14,15].

Determination of UV/RI ratios. After isolation of the polymer, the molecular weight profile was recorded on a GPC equipped with three Waters Ultrastyragel preparatory GPC columns (10^4 Å, 10^3 Å and 500 Å) eluted with tetrahydrofuran, plumbed through a Waters 2487 UV detector, and a Waters 410 RI detector. The UV detector sensitivity was set to 4.00 aufs (absorbance units full scale), and the absorption for all experiments was monitored at 345 nm. The RI sensitivity was set to 8, and the temperature at the detector was kept constant with a thermostat set between 31.9–32.5 °C. After the profile was recorded, the area under the UV and RI curves for unbroken chains was measured using triSEC software (version 3.0, revision B.00.05, Viscotek). The baseline and integration limits were set to include the M_w range corresponding to the unbroken chains using the RI trace. The UV area was then measured without changing the position of the integration limits. The baseline in the UV trace was redrawn to include only the area within the integration limits.

^{13}C NMR experiments. For ^{13}C labelling experiments, the polymer of interest was sonicated as above with ^{13}C-labelled maleimide, **5**^{13}**C**. The polymer from two such runs was isolated by preparatory GPC and combined in an NMR tube with the appropriate solvent. The spectra were recorded on an instrument operating at a ^{13}C frequency of 125 MHz, with about 10–15 mg of polymer dissolved in 0.5 ml of solvent and 15,000–17,000 acquisitions.

Received 29 November 2006; accepted 15 February 2007.

1. Beyer, M. K. & Clausen-Schaumann, H. Mechanochemistry: The mechanical activation of covalent bonds. *Chem. Rev.* **105**, 2921–2948 (2005).
2. Basedow, A. M. & Ebert, K. H. Ultrasonic degradation of polymers in solution. *Adv. Polym. Sci.* **22**, 83–148 (1977).

3. Woodward, R. B. & Hoffmann, R. The conservation of orbital symmetry. *Angew. Chem. Int. Edn Engl.* **8,** 781–853 (1969).
4. Evans, E. Probing the relation between force-lifetime and chemistry in single molecular bonds. *Annu. Rev. Biophys. Biomol. Struct.* **30,** 105–128 (2001).
5. Beyer, M. K. The mechanical strength of a covalent bond calculated by density functional theory. *J. Chem. Phys.* **112,** 7307–7312 (2000).
6. Roth, W. R., Rekowski, V., Börner, S. & Quast, M. Die disrotative cyclobuten-ringöffnung. *Liebigs Ann. Chem.* 409–430 (1996).
7. Boffa, L. S. & Novak, B. M. "Link-functionalized" polymers: an unusual macromolecular architecture through bifunctional initiation. *Macromolecules* **30,** 3494–3506 (1997).
8. Nguyen, T. Q. & Kausch, H.-H. Mechanochemical degradation in transient elongational flow. *Adv. Polym. Sci.* **100,** 73–182 (1992).
9. Berkowski, K. L., Potisek, S. L., Hickenboth, C. R. & Moore, J. S. Ultrasound-induced site-specific cleavage of azo-functionalized poly(ethylene glycol). *Macromolecules* **38,** 8975–8978 (2005).
10. Reddy, P. Y., Kondo, S., Fujita, S. & Toru, T. Efficient synthesis of fluorophore-linked maleimide derivatives. *Synthesis* 999–1002 (1998).
11. Price, G. J. The use of ultrasound for the controlled degradation of polymer solutions. *Adv. Sonochem.* **1,** 231–287 (1990).
12. Arnold, B. J., Sammes, P. G. & Wallace, T. W. Photochemical reactions. Part IV. Thermal generation of photoenols and their derivatives from disubstituted 1,2-dihydrobenzocyclobutenes. *J. Chem. Soc. Perkin Trans. I* 415 (1974).
13. Suslick, K. S., Goodale, J. W., Schubert, P. F. & Wang, H. H. Sonochemistry and sonocatalysis of metal carbonyls. *J. Am. Chem. Soc.* **105,** 5781–5785 (1983).
14. Graessley, W. W. Polymer chain dimensions and the dependence of viscoelastic properties on concentration, molecular weight and solvent power. *Polymer* **21,** 258–262 (1980).
15. Rodd, L. E., Scott, T. P., Boger, D. V., Cooper-White, J. J. & McKinley, G. H. The inertio-elastic planar entry flow of low-viscosity elastic fluids in micro-fabricated geometries. *J. Non-Newtonian Fluid Mech.* **129,** 1–22 (2005).
16. Kersey, F. R., Yount, W. C. & Craig, S. L. Single-molecule force spectroscopy of bimolecular reactions: system homology in the mechanical activation of ligand substitution reactions. *J. Am. Chem. Soc.* **128,** 3886–3887 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Martínez, K. Suslick and P. Beak for discussions. This work was supported by grants from the Air Force Office of Scientific Research and The Petroleum Research Fund.

Author Contributions J.S.M. conceived the BCB experiment. C.R.H. performed all experiments and initial computational analyses. J.B. performed the final computational analysis. S.R. Wilson performed the crystallography. N.R.S., S.R. White and J.S.M. directed the research and all authors wrote the manuscript.

Author Information The atomic coordinates for *cis*-1,2-bis [(3-carboxypropanoyl)oxy]-1,2-dihydrobenzocyclobutene **2** (accession number 628670), *N*-(1-pyrene)-2,3-naphthimide (accession number 628377), and 4,9-diacetoxy-2-(1-pyrenyl)-(3aR,4c,9c,9ac)-3a,4,9,9a-tetrahydro-benzo [f]isoindole-1,3-dione **6** (accession number 628671) have been deposited in the Cambridge Crystal Structure Database. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.S.M. (jsmoore@uiuc.edu).

LETTERS

Seismic evidence for convection-driven motion of the North American plate

David W. Eaton[1] & Andrew Frederiksen[2]

Since the discovery of plate tectonics, the relative importance of driving forces of plate motion has been debated[1,2]. Resolution of this issue has been hindered by uncertainties in estimates of basal traction, which controls the coupling between lithospheric plates and underlying mantle convection[2–4]. Hotspot tracks preserve records of past plate motion[5] and provide markers with which the relative motion between a plate's surface and underlying mantle regions may be examined. Here we show that the 115–140-Myr surface expression of the Great Meteor hotspot track in eastern North America is misaligned with respect to its location at 200 km depth, as inferred from plate-reconstruction models and seismic tomographic studies[6]. The misalignment increases with age and is consistent with westward displacement of the base of the plate relative to its surface, at an average rate of 3.8 ± 1.8 mm yr^{-1}. Here age-constrained 'piercing points' have enabled direct estimation of relative motion between the surface and underside of a plate. The relative displacement of the base is approximately parallel to seismic fast axes and calculated mantle flow[7], suggesting that asthenospheric flow may be deforming the lithospheric keel and exerting a driving force on this part of the North American plate.

Theoretical studies indicate that plate motion is primarily controlled (~90%) by convective flow driven by density heterogeneities in the mantle, particularly those associated with sinking oceanic slabs[1,7–9]. The nature and strength of viscous coupling of tectonic plates to mantle convection remains unclear, however. In western North America where lithosphere is thin (<100 km), indirect evidence from seismic anisotropy suggests weak coupling of plate motion to the underlying flow, resulting in a small drag force[10]. On the other hand, tectonically quiescent continental regions (cratons) are generally underlain by refractory lithospheric keels that may extend 200 km or more into the mantle[11]. Although it is recognized that these keels could increase basal traction[1–4], quantifying this force is hindered by a lack of direct observations[10].

Bokelmann[2,12] proposed that analysis of deformation fabrics of lithospheric keels inferred from seismic anisotropy may provide clues about whether basal traction constitutes a net drive or drag force in these regions (Fig. 1). We chose to study the North American plate because it is among the fastest-moving plates that possess a thick lithospheric keel[3]. Seismic P-wave anisotropy beneath the craton indicates a consistently southwest-dipping deformation fabric, roughly parallel to plate motion, suggesting that North American plate motion is driven (rather than impeded) by interaction of mantle flow with the keel[2,12].

Here we apply similar deformation analysis to an ancient hotspot track. Hotspots represent localized sublithospheric sources of magmatism that exhibit far less motion with respect to a whole-mantle reference frame than most plates[5]. Although hotspots are traditionally interpreted as thermal plumes that originate in the lower mantle[5], alternative models for hotspots include secondary plumes, convective instabilities and fertility variations in the upper mantle[13]. Regardless of source, hotspot-related magmatism enables the determination of relative motion between a plate surface and the underlying mantle.

The Great Meteor hotspot track in the Atlantic and North America has been studied extensively[14–16]. Various crustal features are interpreted to mark the continental segment, delineating an arcuate track (Fig. 2). Igneous crystallization ages increase monotonically along the track (Fig. 2), generally consistent with the hotspot model[17]. Surface features attributed to the hotspot include small-volume kimberlite eruptions that penetrate through thick cratonic lithosphere[17], intermediate-volume alkaline magmas near the edge of the craton[16], and an intervening region of unusually deep crustal seismicity[18]. These changes in near-surface expression may reflect hotspot interaction with progressively thinner lithosphere, owing to motion of the overriding plate (Fig. 1a).

Surface-wave tomographic studies of North America[6] reveal an elongate low-velocity anomaly beneath the Great Lakes. This feature is strongly expressed at 200 km depth (Fig. 2), placing it at (or near) the base of the lithosphere[19]. The anomaly has been interpreted as an indentation, or 'divot', in the high-velocity mantle keel[20]. Curiously, the axis of the divot is better aligned with the oceanic hotspot track (New England seamount chain) than are surface features inferred to mark the continental segment. Here, we explore whether the divot may have formed as a result of the passage of North America over the hotspot.

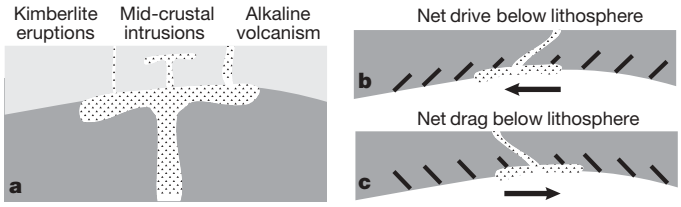


Figure 1 | Schematic diagram showing near-surface manifestations of a hotspot and shear sense from different polarities of basal traction. **a**, Depending on the thickness of the plate, the near-surface expressions of a hotspot may include small-volume kimberlite eruptions that penetrate the lithosphere, intermediate-volume alkaline volcanism and intervening regions where no surface volcanic rocks are evident, but where mid-crustal seismicity is observed. In the latter case, the absence of surface volcanism suggests that intrusions do not penetrate beyond the mid-crust. **b**, If sublithospheric mantle flow exerts a net drive on plate motion, the base of a hotspot track will tend to be displaced in the direction of plate motion, relative to the surface. Deformation fabric and corresponding seismic fast axes are shown schematically by dipping lines[2,12]. **c**, If basal traction exerts a net drag, the opposite sense of shear is produced. In both **b** and **c**, plate-edge forces such as collision resistance contribute to overall torque balance[1]. Creep processes in the mantle may accommodate the postulated deformation of the lithosphere[3].

[1]Department of Earth Sciences, University of Western Ontario, London, Ontario N6A 5B7, Canada. [2]Department of Geological Sciences, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada.

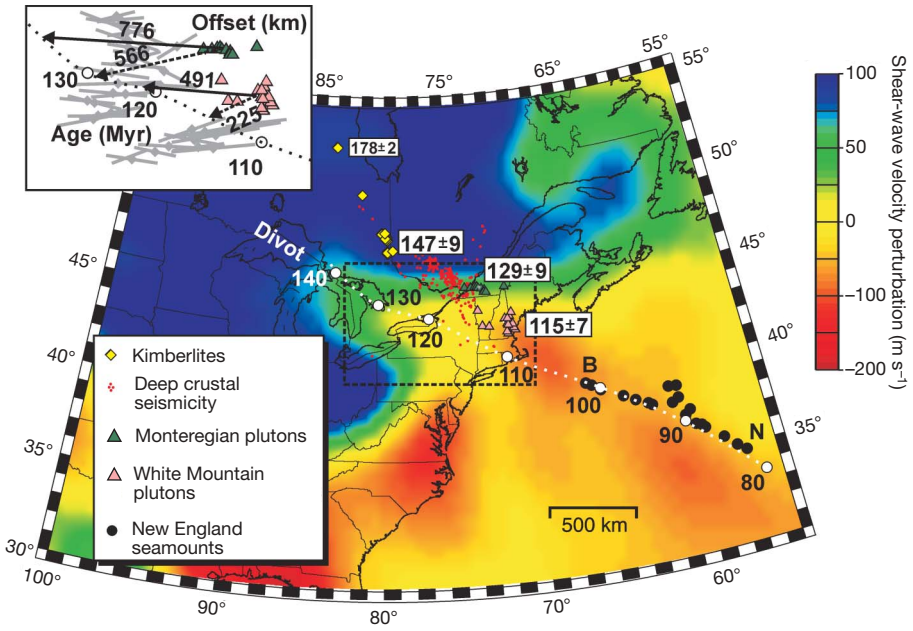


Figure 2 | Inferred track of the Great Meteor hotspot. Track is superimposed on shear-wave velocity perturbation at a depth of 200 km from model NA04 (ref. 6). Blue colours in the map indicate high-velocity regions associated with the mantle keel beneath North America. Numbers inside white rectangles are ages in Myr (mean ± standard deviations) of surface features along the track[17,27,30]. The dashed white line shows the projected position of the Great Meteor hotspot (present location 27° E, 26° N) based on a recent plate-reconstruction model[24]. The projected track has been rotated by 27.2° about a pole at 138.8° E, 13.4° S to minimize the least-squares misfit to the Bear (B) and Nashville (N) seamounts (103 and 82 Myr, respectively). It agrees well with the position of an elongate low-velocity anomaly in the Great Lakes region ('divot'), but is misaligned with the surface expression of the track. The upper inset shows the offset of the centroid location of the Monteregian and White Mountain igneous provinces, relative to coeval points on the projected track (dotted line with circles). Black arrows show displacement for maximum (solid) and mean (dashed) emplacement ages; grey bars show seismic fast-axis directions[29]. The inset region is indicated by the dashed rectangle in the main figure.

Seismic velocity anomalies in the continental upper mantle can be classified as thermal or compositional in origin, or a combination of both[21]. To test whether a thermal anomaly could persist since passage over the hotspot (120 Myr before), we have conducted a two-dimensional numerical experiment. The initial thermal anomaly is represented by a 300 km × 50 km rectangular region (underplate) in the lowermost 50 km of the lithosphere (150–200 km), fed by a 40-km-wide vertical conduit. This region is initially perturbed from a cratonic geotherm by resetting the temperature to adiabatic conditions. For a relative plate velocity of 3 cm yr^{-1} and buoyancy flux of 2,000 kg s^{-1} (ref. 16), 10% of the volume of this zone is replenished continuously by the hotspot. The track is generated much faster than it decays by thermal diffusion, so our two-dimensional treatment is justified. We obtained a finite-difference solution to the diffusion equation using thermal conductivity parameters for this region[22]. After 120 Myr, the residual thermal anomaly is about 80 °C and is most intense at ~200 km depth (Fig. 3). The thermal anomaly is similar in size and shape to the observed seismic anomaly

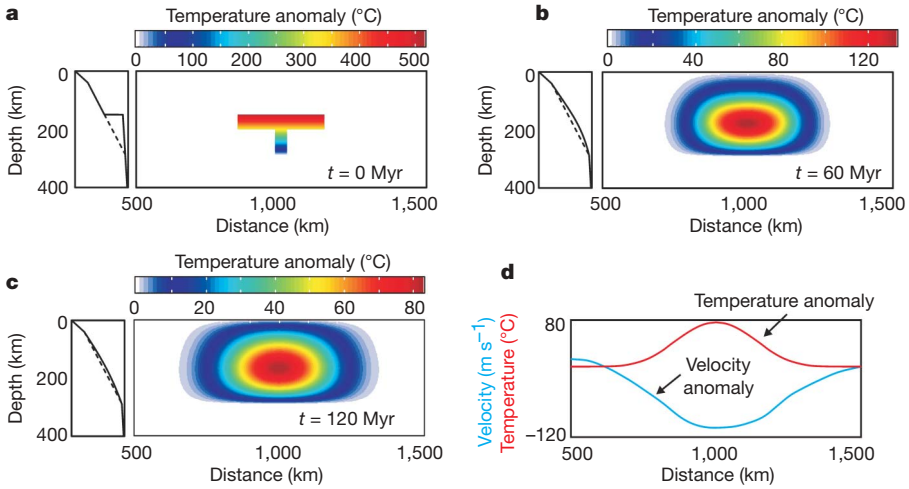


Figure 3 | Modelled evolution of the thermal anomaly from a hotspot. **a**, Passage of a plume at $t = 0$, 'resets' the geotherm by increasing the temperature to the adiabat within a 40-km-wide plume and a 300 km × 50 km underplate region. The graphs on the left show the initial geotherm (dashed) and temperature field (solid) in the centre of the thermal anomaly. The temperature range is 0 °C to 1,500 °C. In **b** and **c**, the subsequent temperatures evolve by diffusion in two dimensions, thus effectively treating the anomaly as infinite and out of the plane. Panel **d** compares the modelled temperature anomaly ($t = 120$ Myr) with the shear-wave velocity anomaly at 200 km depth from model NA04 (ref 6). The shear-wave velocity profile extends from 39.69° N, 83.36° W to 48.09° N, 76.62° W.

in the divot. We note, however, that if the seismic anomaly is entirely thermal in origin, then a thermal sensitivity of $1.1\,m\,s^{-1}\,^{\circ}C^{-1}$ is required, at the upper limit of laboratory-derived values for mantle rocks under subsolidus conditions[21].

It is likely that a mantle anomaly produced by a plume would also have a compositional signature owing to enrichment of the lithospheric keel by plume material[23]. Such chemical enrichment tends to reduce seismic velocity, while increasing density. For mantle rocks an excess temperature of ~80 °C predicted by our model would result in a density decrease of about 0.25% (ref. 23). Isostatic compensation for this mass deficiency would require an increase of 4% in the FeO/(FeO + MgO) ratio, well within the natural range of values for continental lherzolites[23].

We computed the track of the Great Meteor hotspot in North America based on a recent plate reconstruction model[24], constrained to fit the New England seamount trend (Fig. 2). The projected track agrees well with the mantle divot at 200 km depth, but is misaligned with near-surface continental features. Such a discrepancy could arise from deflection of a plume tail by cross-currents in the mantle[5] (the 'mantle wind'). Alternatively, a plume may be deflected by the cratonic keel, in which case its surface expression may follow the path of least mechanical resistance through the lithosphere[25]. Finally, a hotspot may not be stationary with respect to the overall mantle reference frame[26], in which case a projection of the track based on an independently derived plate-reconstruction model may be in error.

If the mantle divot were produced by the Great Meteor hotspot, none of these explanations is consistent with the observed misalignment. Plume deflection by mantle wind is unlikely, because the hotspot track extends sufficiently far into the continent for thick lithosphere west of the track to have shielded it from cross-currents above 200 km. Furthermore, deflection by a dipping zone of weakness does not satisfactorily explain the observed increase in offset with age along the track (see below). Similarly, although a departure from hotspot fixity may explain discrepancies between computed and observed tracks, it does not explain a systematic misalignment between deep and shallow segments.

We propose that the near-surface track of the hotspot was originally emplaced on top of the mantle seismic anomaly, but has since become offset. By connecting coeval points along the inferred shallow and deep trends, we can estimate the net displacement. There is evidence for extended magma residence time in the shallow lithosphere[27], so we assume, for a particular igneous suite, that passage of the hotspot occurred within one standard deviation above the mean age. Using the centroid location (average location of dated intrusions), we estimate a net offset of 556–776 km for the Monteregian intrusions (129 ± 9 Myr) and 225–491 km for the White Mountain intrusions (115 ± 7 Myr) (Fig. 2, inset). Assuming that displacement between the surface and 200 km depth occurred as a continuous process, these estimates indicate an average offset rate of $3.8 \pm 1.8\,mm\,yr^{-1}$, implying a depth-averaged lithospheric strain rate of $6.0 \pm 2.9 \times 10^{-16}\,s^{-1}$.

Instantaneous flow calculations[7] predict subhorizontal westward mantle flow of $\sim 2\,cm\,yr^{-1}$ beneath this part of North America, driven by excess density in the lower mantle from the remnant Farallon slab. Our inferred displacement rate is intermediate between zero (rigid plate) and the calculated asthenospheric flow velocity. We interpret the observed displacement to arise from lithospheric deformation. This interpretation is supported by recent seismic tomographic data, which reveal a southwest-dipping low-velocity anomaly extending upward from 200 km (within the divot) towards the surface near the Monteregian intrusions[28].

The deformation implied by our model represents horizontal simple shear in the lithospheric mantle keel arising from viscous coupling with underlying asthenospheric flow, for which strain compatibility can be maintained in the mantle independently of crust and surface deformation[3]. The sense of shear supports Bokelmann's hypothesis[2,12] that mantle flow is deforming the cratonic keel and thus in part driving the motion of North America (Fig. 1b). Assuming a simple newtonian rheology, for an effective viscosity of $\sim 5 \times 10^{21}$ Pa s (ref. 3) our estimated strain rate of $\sim 6.0 \times 10^{-16}\,s^{-1}$ implies a basal shear stress of ~3 MPa. This stress magnitude is of the same order as recent estimates of basal traction beneath thick lithospheric roots, on the basis of laboratory data[3] and mantle-flow modelling[4].

Seismic anisotropy induced by basal traction should lead to shear-wave fast-polarization directions that are parallel, to within 180° uncertainty, to the lithosphere–asthenosphere velocity difference[10]. In many continental regions, such a relationship is obscured by multiple layers of anisotropy (for example, ref. 3), but in the region of the mantle divot only a single anisotropic layer is observed[29]. Both asthenospheric flow[7] and absolute plate motion[29] are approximately westward in this area, so flow-induced anisotropy from basal traction should result in a simple pattern of east–west fast-splitting directions. Shear-wave splitting results (Fig. 2, inset) confirm this orientation, providing further support for our model.

Thus our study documents a conspicuous misalignment between an elongate seismic velocity anomaly at 200 km depth (the so-called Great Lakes mantle divot) and the surface track of the Great Meteor hotspot. The mantle divot is compatible with the projected track of the hotspot and may be caused by the combined effects of residual thermal perturbation and compositional changes. We propose that the base of the track is displaced westward from the surface as a result of viscous coupling of the North American cratonic keel with mantle flow. Our observations represent the first instance, to our knowledge, in which age-constrained piercing points are available to estimate net displacement between the top and base of a continental plate. The inferred sense of shear implies that basal traction may in part drive plate motion.

Received 23 June 2006; accepted 9 February 2007.

1. Forsyth, D. & Uyeda, S. On the relative importance of the driving forces of plate motion. *Geophys. J. R. Astron. Soc.* **43,** 163–200 (1975).
2. Bokelmann, G. H. R. Which forces drive North America? *Geology* **30,** 1027–1030 (2002).
3. Bokelmann, G. H. R. & Silver, P. Shear stress at the base of shield lithosphere. *Geophys. Res. Lett.* **29,** doi:10.1029/2002GL015925 (2002).
4. Conrad, C. P. & Lithgow-Bertelloni, C. Influence of continental roots and asthenosphere on plate-mantle coupling. *Geophys. Res. Lett.* **33,** doi:10.1029/2005GL025621 (2006).
5. Duncan, R. A. & Richards, M. A. Hotspots, mantle plumes, flood basalts, and true polar wander. *Rev. Geophys.* **29,** 31–50 (1991).
6. Van der Lee, S. & Frederiksen, A. in *Seismic Data Analysis and Imaging With Global and Local Arrays* (eds Nolet, G. & Levander, A.) 67–80 (AGU Geophysical Monograph 157, Washington DC, 2005).
7. Conrad, C. P. & Lithgow-Bertelloni, C. Iceland, the Farallon slab, and dynamic topography of the North Atlantic. *Geology* **32,** 177–180 (2004).
8. Lithgow-Bertelloni, C. L. & Richards, M. A. The dynamics of Cenozoic and Mesozoic plate motions. *Rev. Geophys.* **36,** 27–78 (1998).
9. Conrad, C. P. & Lithgow-Bertelloni, C. How mantle slabs drive plate tectonics. *Science* **298,** 207–209 (2002).
10. Silver, P. G. & Holt, W. E. The mantle flow field beneath western North America. *Science* **295,** 1054–1057 (2002).
11. Jordan, T. H. The continental tectosphere. *Rev. Geophys.* **13,** 1–12 (1975).
12. Bokelmann, G. H. R. Convection-driven motion of the North American craton: Evidence from P-wave anisotropy. *Geophys. J. Int.* **148,** 278–287 (2002).
13. Courtillot, V., Davaille, A., Besse, J. & Stock, J. Three distinct types of hotspots in the Earth's mantle. *Earth Planet. Sci. Lett.* **205,** 295–308 (2003).
14. Crough, S. T. Mesozoic hotspot epeirogeny in eastern North America. *Geology* **9,** 2–6 (1981).
15. Morgan, W. J. Hotspot tracks and the early rifting of the Atlantic. *Tectonophysics* **94,** 123–139 (1983).
16. Sleep, N. H. Monteregian hotspot track: A long-lived mantle plume. *J. Geophys. Res.* **95,** 21983–21990 (1990).
17. Heaman, L. M. & Kjarsgaard, B. A. Timing of eastern North American kimberlite magmatism: continental extension of the Great Meteor hotspot track? *Earth Planet. Sci. Lett.* **178,** 253–268 (2000).
18. Adams, J. & Basham, P. in *Neotectonics of North America* (eds Slemmons, D. B., Engdahl, E. R., Zoback, M. D. & Blackwell, D. D.) 261–276 (Geological Society of America, Boulder, Colorado, 1991).
19. Darbyshire, F. A., Eaton, D. W., Frederiksen, A. W. & Ertolahti, L. New insights into the lithosphere of the Superior Province from Rayleigh wave dispersion and receiver function analysis. *Geophys. J. Int.* (in the press).

20. Fouch, M. J., Fischer, K. M., Parmentier, E. M., Wysession, M. E. & Clarke, T. J. Shear wave splitting, continental keels, and patterns of mantle flow. *J. Geophys. Res.* **105,** 6255–6275 (2000).
21. Artemieva, I. M., Magali, B., Leveque, J.-J. & Mooney, W. D. Shear wave velocity, seismic attenuation, and thermal structure of the continental upper mantle. *Geophys. J. Int.* **157,** 607–628 (2004).
22. Mareschal, J.-C. *et al.* Heat flow and deep thermal structure near the southeastern edge of the Canadian Shield. *Can. J. Earth Sci.* **37,** 399–414 (2000).
23. VanDecar, J. C., James, D. E. & Assumpção, M. Seismic evidence for a fossil mantle plume beneath South America and implications for plate driving forces. *Nature* **378,** 25–31 (1995).
24. Schettino, A. & Scotese, C. R. Apparent polar wander paths for the major continents (200 Ma to the present day): A palaeomagnetic reference frame for global plate tectonic reconstructions. *Geophys. J. Int.* **163,** 727–759 (2005).
25. Sleep, N. H., Ebinger, C. J. & Kendall, J.-M. in *The Early Earth: Physical, Chemical and Biological Development* (eds Fowler, C. M. R., Ebinger, C. J. & Hawkesworth, C. J.) 135–150 (GSL Special Publication 199, Geological Society of London, London, 2002).
26. Tarduno, J. A. & Gee, J. Large-scale motion between Pacific and Atlantic hotspots. *Nature* **378,** 477–480 (1995).
27. Eby, G. N. Geochronology of the Monteregian Hills alkaline igneous province, Quebec. *Geology* **12,** 468–470 (1984).
28. Aktas, K. & Eaton, D. W. Upper-mantle velocity structure of the lower Great Lakes region. *Tectonophysics* **420,** 267–281 (2006).
29. Eaton, D. W., Frederiksen, A. & Miong, S.-Y. Shear-wave splitting observations in the lower Great Lakes region: Evidence for regional anisotropic domains and keel-modified asthenospheric flow. *Geophys. Res. Lett.* **31,** L07610, doi:10.1029/2004GL019438 (2004).
30. Zartman, R. E. Geochronology of some alkaline rock provinces in eastern and central United States. *Annu. Rev. Earth Planet. Sci.* **5,** 257–286 (1977).

Acknowledgements This work was supported by NSERC Discovery grants and benefited from discussions with J. Adams, D. Moser, P. McCausland, D. Jiang and F. Darbyshire. A review by M. Savage improved the clarity of this manuscript.

Author Contributions D.W.E. performed thermal modelling and data analysis; A.F. provided seismic tomography results and interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.W.E. (deaton@uwo.ca).

LETTERS

The effects of competition and predation on diversification in a model adaptive radiation

Justin R. Meyer[1] & Rees Kassen[1]

Much of life's diversity is thought to have arisen through successive rounds of adaptive radiation—the rapid diversification of a lineage into a range of ecologically and phenotypically distinct species[1–3]. Both resource competition and predation have been suggested as mechanisms driving this process[4,5], although the former is better studied than the latter[6,7]. Here we show experimentally how predation by a protist, *Tetrahymena thermophila*, affects diversification in a model adaptive radiation of the bacterial prey, *Pseudomonas fluorescens*. We estimate the frequency-dependent fitness functions of competing niche-specialist prey in the presence and absence of predation, and use these to test hypotheses about the extent (measured as the number of new genotypes) and rate of diversification. Competition and predation independently generated diversifying selection that we show is capable of driving prey diversification to similar extents but at different rates, diversification being markedly delayed in the presence of predators. The cause of this delay stems from weaker diversifying selection due to the reduction in prey density caused by predation. Our results suggest that predation may play an under-appreciated role in driving adaptive radiations.

The rapid diversification characteristic of adaptive radiation is thought to be driven by strong divergent natural selection generated through resource competition in the presence of abundant ecological opportunity (defined as vacant niche space)[4]. The impact of predation on the extent and rate of diversification in prey remains controversial[6]. Comparative[8,9] and experimental[10] evidence clearly implicates a role for predation in generating novel ecological opportunities through the evolution of predator-resistance strategies or access to predator-free space. The effect of predation on the strength of diversifying selection, and so both the extent of phenotypic divergence and the rate of diversification, is less clear. Predation may interact synergistically with competition to promote diversification if, for example, phenotypically intermediate types are susceptible to predation[11,12]. Alternatively, diversification may be slowed, or halted altogether, if predators reduce prey densities sufficiently to prevent resource competition[13].

To understand the manifold effects of predation on diversification, we estimated the strength of diversifying selection due to competition, predation and their combination by measuring the frequency-dependent fitness functions of two of the most common niche-specialist genotypes observed during the model adaptive radiation of *Pseudomonas fluorescens* SBW25 in spatially structured (static) microcosms. The two genotypes are the ancestral broth-colonizing 'smooth' and the biofilm-forming 'wrinkly spreader'. Wrinkly spreader, which occupies the air–broth interface of a static microcosm, is invariably the first niche-specialist to arise by mutation during diversification from the ancestral smooth[14] and in the absence of predation both types are stably maintained by negative frequency-dependent selection in static microcosms[15]. The slope of the line relating wrinkly spreader relative fitness to its initial frequency gives an estimate of the strength of frequency-dependent selection; a negative slope with an x-intercept between zero and one indicates that selection is divergent, each type being fittest when rare. The slope of this relationship in the absence and presence of predators provides estimates of the strength of divergent selection due to competition and competition in combination with predation, respectively. We obtained estimates of the effect of predation independently of competition by adding an antibiotic to the medium at a concentration that prevents bacterial growth without killing the cells and to which the predator is resistant (see Methods).

Our results, shown in Fig. 1, reveal two notable effects of predation on frequency-dependent selection. First, fitness is always negatively frequency-dependent (slope ± s.e. tested using a two-tailed, one-way t-test against a null hypothesis of zero slope; competition alone: -8.74 ± 0.95, $t = -9.22$, d.f. $= 53$, $P < 0.0001$; predation alone: -4.66 ± 0.69, $t = -6.79$, d.f. $= 57$, $P < 0.0001$; competition and predation: -3.35 ± 0.96, $t = -3.47$, d.f. $= 48$, $P < 0.0001$), suggesting the operation of diversifying natural selection in both the presence and absence of predation. Second, predation reduces the strength of frequency-dependent selection, a result confirmed by the significance of the interaction between the starting frequency of

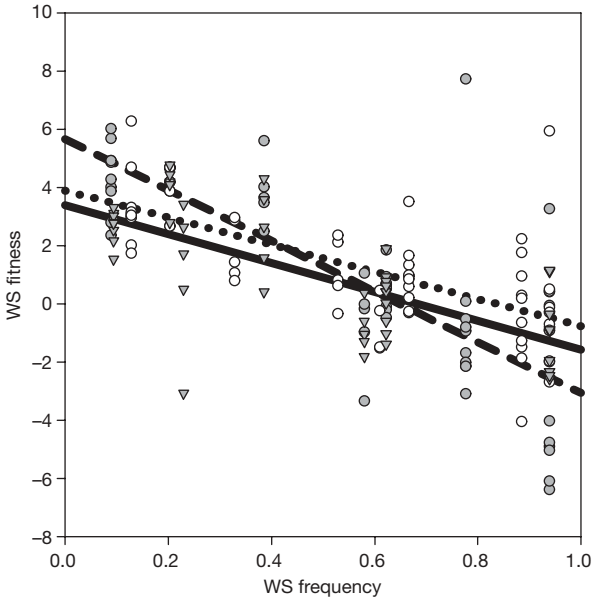


Figure 1 | Frequency-dependent fitness functions due to competition (dashed line and filled circles), predation (dotted line and open circles) and their interaction (solid line and triangles). Fitness is measured as the selection rate per day (r), defined as the difference in the malthusian parameters between wrinkly spreader (WS) and smooth genotypes. Frequency is an estimate of initial wrinkly spreader frequency.

[1]Department of Biology and Center for Advanced Research in Environmental Genomics, University of Ottawa, Ottawa K1N 6N5, Canada.

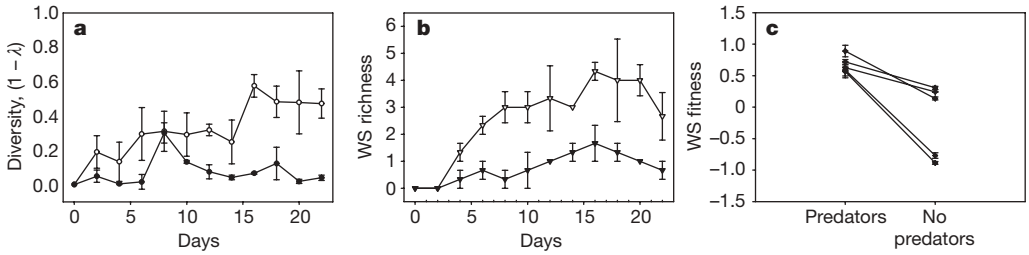


Figure 2 | The effect of predation on the diversification of wrinkly spreaders in minimal media. **a**, Diversity, measured as the complement of Simpson's index ($1-\lambda$), is greater in the presence of predators (open circles) than in their absence (filled circles). **b**, The increased diversity is due to the emergence of additional wrinkly spreader genotypes, WS richness, with predators (open triangles) than without predators (filled triangles). **c**, Five phenotypically distinct wrinkly spreader genotypes had increased fitness in the presence of predators compared to that in their absence. Error bars, ± 1 s.e.m. ($n = 3$).

wrinkly spreader and treatment in an analysis of covariance ($F_{2,158} = 10.33$, $P < 0.0001$).

The observation of negative frequency-dependent fitness between smooth and wrinkly spreader in the absence of predators is consistent with previous results[15], and probably stems from the intense competition for limiting resources, especially oxygen, caused by the combination of spatial structure and population growth. Wrinkly spreader has a fitness advantage when rare because it forms a self-supporting mat that colonizes the air–liquid interface and permits access to oxygen. Fitness declines as it becomes more common, both because resources are relatively more abundant in the broth and because the mat becomes heavier, eventually sinking under its own weight[15,16]. In the presence of *T. thermophila*, who are generalist filter-feeders, selection remains negatively frequency-dependent, as before, but its strength is weakened, probably because resource competition is weakened by the approximately tenfold reduction in prey density due to predation (bacterial cell density per ml $\pm$ 1 s.e. as follows: without predators, $(2.86 \pm 0.65) \times 10^9$; with predators, $(2.78 \pm 0.79) \times 10^8$; ANOVA: $F_{1,10} = 38.22$, $P < 0.0001$). Fitness remains frequency-dependent because the mat provides a refuge from predators for wrinkly spreaders when rare (as in ref. 17). We suspect that wrinkly-spreader fitness declines when common in the presence of predators, however, because *T. thermophila* preferentially feeds at the surface where oxygen is more abundant, an observation we have confirmed by contrasting the abundance of *T. thermophila* just below the air–broth interface in static microcosms with and without mineral oil on the surface, which acts to restrict the diffusion of oxygen into the media (mean density of *T. thermophila* per ml $\pm$ s.e. as follows: with oil, 8,888 $\pm$ 320; no oil, 2,777 $\pm$ 320: two-sample *t*-test, $t_4 = 8.61$, $P = 0.0007$; see Methods).

The negative frequency-dependent fitness between smooth and wrinkly spreader that we have observed is a hallmark of diversifying selection associated with adaptive radiation[18]. Moreover, the observation that predation generates diversifying selection independently of competition suggests that predators should be capable of driving diversification in *P. fluorescens* under conditions where resource competition cannot. To test this prediction, we cultured replicate populations of initially isogenic *P. fluorescens* in the presence and absence of predators in static microcosms containing a minimal salts medium with glucose as the sole carbon source. Previous work has shown that the characteristic *P. fluorescens* adaptive radiation does not occur under these conditions, although wrinkly spreaders occasionally arise at low frequency[19]. Our results indicate that diversification was more extensive in the presence of predators than in their absence (Fig. 2a), and that this was largely owing to the emergence of novel wrinkly spreader genotypes (Fig. 2b) resistant to predation (Fig. 2c; analysis of variance contrasting fitness in the presence and absence of predators for five independent wrinkly spreaders isolated from microcosms containing predators: $F = 34.57$, $P < 0.0001$, d.f. = 27). These results suggest that predation alone is sufficient to drive adaptive radiation through the creation of ecological opportunity afforded by resistance.

Predators also decreased the strength of frequency-dependent selection, suggesting that a second effect of predation on adaptive radiation may be to slow the rate of diversification. To test this prediction, we followed the progress of diversification in the presence and absence of the predator by destructively sampling static KB (King's B medium) microcosms inoculated with the smooth ancestor intensively over the course of eight days. The results, shown in Fig. 3, are striking. The dynamics of diversification resemble the characteristic logistic pattern expected from theory[20–22], the major difference between treatments being the timing at which diversity arises: average richness among replicates becomes significantly greater than one after 48 h in the absence of predators (one-way *t*-test: $t = 3.21$, d.f. = 4, $P = 0.0163$) and 64 h in their presence (one-way *t*-test: $t = 2.45$, d.f. = 4, $P = 0.035$). This result is consistent with those from simple models of the dynamics of gene frequencies under different strengths of frequency-dependent selection (see Supplementary Information). Given the large prey population sizes (approximately 10^6–10^7 cells ml^{-1} in the presence of predators after 24 h in this experiment), this delay in diversification is unlikely to be caused by changes in the supply of novel genotypes through mutation. We also observed a slower maximal rate of increase in richness in the presence of predators, though this difference was not significant ($b \pm 95\%$ confidence limits (CL): no predators, 0.102 ± 0.010 for time = 32–104 h; predators, 0.088 ± 0.017 for time = 48–104 h). We could not detect a difference in the average maximum diversity (mean $\pm$ 95% CL between 112 and 200 h in both cases: no predators, 6.24 ± 0.32; predators, 5.50 ± 0.49). Parameter estimates obtained from fitting a logistic growth model having a breakpoint and incorporating a time lag before diversification begins lend further support to these results (see Fig. 3 and Supplementary Information).

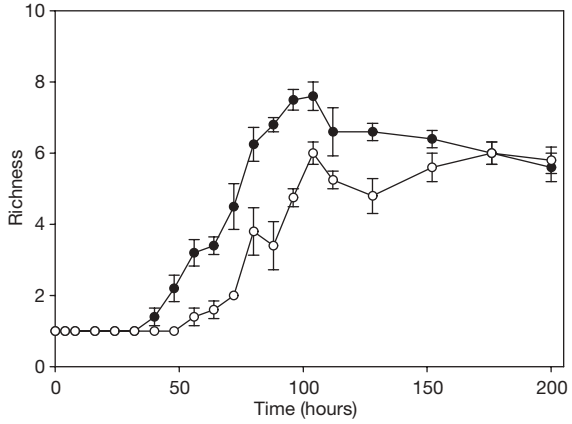


Figure 3 | Dynamics of diversification in the presence and absence of predators. Each circle represents the mean richness of five replicate microcosms, for microcosms with predators (open circles) and without predators (filled circles). Error bars, ± 1 s.e.m.

Our results suggest that predators can modulate the progress of an adaptive radiation in at least two ways. First, predation—by itself or in concert with competition—may generate diversifying selection by creating novel ecological opportunities in the form of access to predator-free space. This suggestion has been made previously[6], and experimental evidence exists to support a role for predation in driving phenotypic divergence among prey[10] and the *de novo* evolution of a single predator-resistant type[23,24]. However, ours is the first direct experimental evidence that predation spurs diversification of prey into a range of predator-resistant phenotypes characteristic of adaptive radiation. Our work also highlights a phenomenon not previously observed: the effect of predation on diversification is most pronounced when the underlying range of resources, and so opportunities for ecological specialization, are limited. Note also that a comparable experiment[13] to ours observed the opposite effect: a bacteriophage predator imposed strong directional selection for resistance, leading to the emergence of a single resistant type whose phenotype differed among microcosms owing to the stochastic effects of low prey population densities caused by infection.

Second, and perhaps less intuitively, predation may slow or delay diversification when the range of available resources is sufficiently large to generate abundant vacant niche space. The underlying cause of this effect stems from changes to the strength of diversifying selection mediated by the combination of resource competition and the selective pressure imposed by predation. Slower rates of diversification due to predation have not previously been observed and may explain, in combination with the greater ecological opportunity afforded by islands, why diversification rates on islands may often be faster than on continents[25].

These results may also help explain why the most spectacular radiations seem to be driven primarily by competition and not predation[26,27]. It seems reasonable to suggest that the range of ecological opportunities offered by resource availability at the lowest trophic levels will often be larger than those generated solely by predation. If so, then predation may only be observed to cause adaptive radiation in exceptionally stringent circumstances, namely, when there are few opportunities for resource specialization. That being said, our observation that predation can spur diversification to similar phenotypic end-points as competition underscores the difficulties faced in teasing apart these two mechanisms in natural systems. The primacy of resource competition in adaptive radiation may be more apparent than real, reflecting the practical challenges associated with conducting experiments capable of disentangling the effects of competition from predation[6] rather than any real bias in the causes of diversification in nature. Either way, our work suggests that predation may play an important—and perhaps under-appreciated—role in explaining the spatially and temporally episodic nature of diversification during the history of life[28].

METHODS

Frequency-dependent fitness experiments. We estimated the fitness of a wrinkly-spreader (WS) genotype chosen at random from a diverse community of *P. fluorescens* that evolved in a KB microcosm, in competition against the ancestral smooth (SM) morph across a range of seven starting frequencies. Fitness was calculated as the selection rate constant r using the equation[29]: $r = (\ln[WS_{final}/WS_{initial}] - \ln[SM_{final}/SM_{initial}])/\text{time}$, where the subscripts final and initial refer to the densities of cells (per ml) of each genotype at the end or beginning of the experiment, respectively. The slope of the regression of fitness against starting frequency gives an estimate of the strength of frequency-dependent selection. Space and time constraints necessitated blocking the experiment according to starting frequency. Densities of the two genotypes were estimated by destructively sampling ten microcosms five hours after inoculation, at which time experimental manipulations, if any, were performed (see below). The remaining forty microcosms were then sampled again after five further hours to obtain estimates of final densities.

We obtained fitness estimates for each treatment—competition alone, predation alone, the interaction between the two, and an antibiotic control—using the following procedure. 1 ml of a defined mixture of smooth and wrinkly-spreader genotypes obtained from pre-inoculation cultures grown overnight under shaken conditions in KB medium was inoculated into static microcosms containing 5 ml of KB at 28 °C. Experimental manipulations were performed five hours following inoculation, which is sufficient time for the initial formation of a biofilm. 1×10^4 cells per ml of *T. thermophila* (strain SB280) were added using a sterile syringe to puncture the biofilm without disturbing the spatial structure of the microcosm. To examine the effect of predation independently of competition, we also added 6.75 μg of the bacteriostatic antibiotic streptomycin just before adding the predator, as well as to control microcosms lacking the predator. By comparing the initial bacterial densities to the control we confirmed that this concentration of streptomycin prevents bacterial growth but does not kill the cells ($F_{1,81} = 0.46$, $P = 0.4998$). Additionally, the mean genotypic densities in the antibiotic treatment were used as the 'initial' densities when calculating r, to compensate for subtle non-significant effects of the antibiotics. Under these conditions, *T. thermophila* is resistant to streptomycin, and visual inspection using a microscope confirmed that protist behaviour remained unaltered. Visual observation of the microcosms also confirmed that the integrity of the biofilms was maintained during manipulation.

The large size of this experiment requires that we restrict attention to the interaction between a single pair of genotypes. Although genetic variation in frequency-dependent fitness among independently evolved wrinkly spreader genotypes has been observed[15,30], it is unlikely to represent a major source of bias here as our main concern was the relative, rather than absolute, effect of predation on frequency-dependent selection. Interestingly, our fitness estimate in the presence of predators predicted well the timing of diversification in an unrelated experiment (Fig. 3 and Supplementary Information), lending further support to the idea that these fitness estimates are unlikely to be severely biased.

Density of *T. thermophila* at the air–broth interface. Six replicate KB microcosms were inoculated with approximately 6×10^4 *T. thermophila* cells followed by the addition of 2 ml of mineral oil to the surface of three of the microcosms. 100 μl of medium was removed from the surface of each microcosm initially to ensure that each test tube was inoculated with the same concentration of *T. thermophila* ($t_4 = 0.0813$, $P = 0.9390$) and again after 4 h of static incubation at 28 °C. Counts were conducted using a haemocytometer.

Diversification in minimal media. Following ref. 15, we tracked the diversification of the ancestral smooth genotype (*P. fluorescens* SBW25) by destructively sampling static microcosms composed of M9 minimal medium supplemented with 3.14 g l^{-1} glucose with or without approximately 1×10^3 predators. Three replicate samples were removed every two days for 22 days. Prey diversity was estimated by plating on KB agar and noting the morphology of at least 100 colonies. *T. thermophila* densities were monitored by vortexing the microcosm for 45 s and counting a sample with a haemocytometer. Diversity was measured either as the complement of Simpson's index, $1 - \lambda = 1 - \sum p_i^2$ where p_i is the frequency of the ith type (following refs 13 and 16) or as richness, the number of morphologically distinct genotypes.

Wrinkly spreader fitness with and without predators. Five independently isolated and phenotypically distinct wrinkly spreader genotypes that had evolved in glucose microcosms with *T. thermophila* were competed separately against the ancestral smooth in minimal medium microcosms containing glucose as the sole source of carbon (as above) with and without predators. We inoculated 5 ml of media with approximately 1×10^4 c.f.u. ml^{-1} of a 1:1 by volume mixture of smooth and wrinkly spreader into microcosms containing 1×10^3 individuals of *T. thermophila* and control microcosms lacking the predator. The experiment was performed in triplicate and sampled after four days. At least 50 colonies of each type were counted to estimate genotypic densities, and the difference in the densities was used to compute the wrinkly spreader:smooth relative fitness (as above).

Rates of diversification. Diversity was estimated as for the minimal medium experiment outlined above but with KB medium and destructively sampling five replicate microcosms from each treatment every eight hours for five days, then once a day for another three days.

Received 22 November 2006; accepted 15 January 2007.

1. Simpson, G. G. *The Major Features of Evolution* (Columbia Univ. Press, New York, 1953).
2. Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381,** 226–229 (1996).
3. Benton, M. J. Diversification and extinction in the history of life. *Science* **268,** 52–58 (1996).
4. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, Oxford, UK, 2000).
5. Van Valen, L. Predation and species diversity. *J. Theor. Biol.* **44,** 19–21 (1974).
6. Vamosi, S. M. On the role of enemies in divergence and diversification of prey: a review and synthesis. *Can. J. Zool.* **83,** 894–910 (2002).
7. Chase, C. M. *et al.* The interaction between predation and competition: a review and synthesis. *Ecol. Lett.* **5,** 302–315 (2002).

8. Welborn, G. A., Shelly, D. K. & Werner, E. E. Mechanisms creating community structure across a freshwater habitat gradient. *Annu. Rev. Ecol. Syst.* **27,** 337–363 (1996).
9. Jablonski, D. & Sepkoski, J. J. Jr. Paleobiology, community ecology, and scales of ecological pattern. *Ecology* **77,** 1367–1378 (1996).
10. Nosil, P. & Crespi, B. J. Experimental evidence that predation promotes divergence in adaptive radiation. *Proc. Natl Acad. Sci. USA* **103,** 9090–9095 (2006).
11. Fryer, G. & Iles, T. D. *The Cichlid Fishes of the Great Lakes of Africa* (Oliver and Boyd, Edinburgh, UK, 1972).
12. Rundle, H. D., Vamosi, S. M. & Schluter, D. Experimental test of predation's effect on divergent selection during character displacement in sticklebacks. *Proc. Natl Acad. Sci. USA* **100,** 14943–14948 (2003).
13. Buckling, A. & Rainey, P. B. The role of parasites in sympatric and allopatric diversification. *Nature* **420,** 496–499 (2002).
14. Goymer, P. *et al.* Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. II. Role of the GGDEF regulator WspR in evolution and development of the wrinkly spreader phenotype. *Genetics* **173,** 515–526 (2006).
15. Rainey, P. B. & Travisano, M. Adaptive radiation in a heterogeneous environment. *Nature* **394,** 69–72 (1998).
16. Kassen, R., Llewellyn, M. & Rainey, P. B. Ecological constraints on diversification in a model adaptive radiation. *Nature* **431,** 984–988 (2004).
17. Hall-Stoodley, L. & Stoodley, P. Biofilm formation and dispersal and the transmission of human pathogens. *Trends Microbiol.* **13,** 7–10 (2005).
18. Clarke, B. C., Partridge, L. & Robertson, A. *Frequency-dependent Selection* (Cambridge Univ. Press, New York, 1988).
19. Travisano, M. & Rainey, P. B. Studies of adaptive radiation using microbial model systems. *Am. Nat.* **156,** S35–S44 (2000).
20. Sepkoski, J. J. A kinetic model of Phanerozoic taxonomic diversity. *Paleobiology* **10,** 246–267 (1984).
21. Walker, T. D. & Valentine, J. W. Equilibrium models of evolutionary species diversity and the number of empty niches. *Am. Nat.* **124,** 887–899 (1984).
22. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, UK, 1995).
23. Bohannan, B. J. M. & Lenski, R. E. Effect of prey heterogeneity on the response of a model food chain to resource enrichment. *Am. Nat.* **153,** 73–82 (1999).
24. Schrag, S. J. & Mittler, J. E. Host-parasite coexistence: The role of spatial refuges in stabilizing bacteria-phage interactions. *Am. Nat.* **148,** 348–377 (1996).
25. Grant, P. R. *Evolution on Islands* (Oxford Univ. Press, Oxford, UK, 1998).
26. Lack, D. *Darwin's Finches* (Cambridge Univ. Press, Cambridge, UK, 1947).
27. Schluter, D. Experimental evidence that competition promotes divergence in adaptive radiation. *Science* **266,** 798–801 (1994).
28. Jablonski, D. The future of the fossil record. *Science* **284,** 2114–2116 (1999).
29. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am. Nat.* **138,** 1315–1341 (1991).
30. MacLean, C. R., Bell, G. & Rainey, P. B. The evolution of pleiotropic fitness tradeoff in *Pseudomonas fluorescens*. *Proc. Natl Acad. Sci. USA* **101,** 8072–8077 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Remold for advice on statistics, P. Rainey for experimental guidance, N. Hairston Jr and H. Rundle for conceptual help, E. Orias and D. Cassidy-Hanely for *T. thermophila* strains, T. Uy for algebraic derivations, and G. Schroeder, J.-N. Jasmin, T. Boland and M. Al-Azzabi for assistance in the laboratory. T. Fukami and L. Harmon provided comments on a draft of the manuscript. This work was supported by an NSERC (Canada) Discovery Grant (to R.K.).

Author Contributions J.R.M. designed and performed all experiments and wrote the paper; R.K. discussed the research, performed statistical analyses, and helped with the writing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.R.M. (justin.raymond.meyer@gmail.com).

LETTERS

Immigration history controls diversification in experimental adaptive radiation

Tadashi Fukami[1], Hubertus J. E. Beaumont[2], Xue-Xian Zhang[2] & Paul B. Rainey[2]

Diversity in biological communities is a historical product of immigration, diversification and extinction[1–4], but the combined effect of these processes is poorly understood. Here we show that the order and timing of immigration controls the extent of diversification. When an ancestral bacterial genotype was introduced into a spatially structured habitat, it rapidly diversified into multiple niche-specialist types[5]. However, diversification was suppressed when a niche-specialist type was introduced before, or shortly after, introduction of the ancestral genotype. In contrast, little suppression occurred when the same niche specialist was introduced relatively late. The negative impact of early arriving immigrants was attributable to the historically sensitive outcome of interactions involving neutral competition[3] and indirect facilitation. Ultimately, the entire boom-and-bust dynamics of adaptive radiation were altered. These results demonstrate that immigration and diversification are tightly linked processes, with small differences in immigration history greatly affecting the evolutionary emergence of diversity.

MacArthur and Wilson's theory of island biogeography explains species diversity as a balance between immigration rate and extinction rate[6]. Although celebrated for simplicity and generality, this model leaves much variation in diversity unexplained[7]. Over recent years, two extensions to the basic theory have emerged: one recognizes that diversity is influenced not just by immigration rate, but also by immigration history (the order and timing of species arrivals)[8–11]; the other stresses the role of diversification (the evolutionary process of lineage splitting)[2–4,12]. Here we explore the interplay between immigration history and diversification.

Whereas historical contingency in evolution has been investigated[13–15], few studies have considered evolutionary effects of immigration history. For example, adaptive radiation has often been assumed to occur with the immigration of a single ancestral species[16]. In reality, immigration of multiple species in various orders and timings is possible[17,18]. Consider an ancestral species that colonizes an island from a distant mainland and subsequently diversifies into multiple species. It is likely that nearby islands will receive multiple species, both ancestral and derived, and that immigration history will vary among islands. Currently we know little about the implications that such variation in immigration has for diversity over evolutionary time. Some insight has arisen from phylogenetic analyses of community assembly[16,19,20], but these approaches provide little detailed information on the historical effects of immigration[18,20,21]. An ideal study would experimentally manipulate the timing of immigration events and observe directly the consequences for the evolutionary emergence of diversity.

Such experiments are difficult to perform on most organisms, but uniquely feasible with bacteria. In particular, experimental populations of *Pseudomonas fluorescens* SBW25 have been used to study adaptive radiation[5,22,23]. *P. fluorescens* populations founded from a single ancestral genotype diversify rapidly when introduced into a spatially structured aquatic microcosm[5]. Three main classes of genetically determined niche-specialist genotypes arise: the smooth morph class (SM), which resembles the ancestral type and primarily colonizes the liquid phase; the wrinkly-spreader morph class (WS), which forms biofilms at the air–liquid interface; and the fuzzy-spreader morph class (FS), which inhabits the microcosm bottom. These types are readily distinguishable by colony morphology[5], and considerable heritable variation exists within each class, particularly within WS (ref. 24). Moreover, it is easy to isolate and store these genotypes, and to regulate the rate at which they migrate to new microcosms. Because reproduction in *P. fluorescens* is entirely asexual, these genotypes are analogous to species in other organisms[22].

To investigate the effect of immigration history on the extent of diversification, the ancestral SM genotype (source SM population) and a derived WS genotype were introduced into spatially structured microcosms in different orders and at different intervals of time (see Methods). This design simulates natural situations in which a source population of a single ecological type colonizes an island of an archipelago and diversifies on that island and, subsequently, the source population and the new populations derived from it emigrate to nearby islands. In such situations, stochastic forces influencing the timing of dispersal make immigration history variable among islands. Our experiment included a control treatment in which the source SM was introduced by itself to new microcosms (see Methods). Because each of the immigrant populations is isogenic, any genotypic variation that arises after introduction can be attributed to evolutionary diversification by *de novo* mutation from the ancestral genotype. Furthermore, the use of a neutral genetic marker (see Methods) enables the origin of populations of each new genotype to be determined.

When introduced alone, the source SM population rapidly diversified into a FS population and multiple WS populations (Fig. 1). Four WS genotypes were distinguishable on the basis of heritable differences in colony morphology, which we refer to as small-WS, large-WS, wheel-WS and SM-like-WS. All new genotypes emerged by day four and coexisted with the source SM population for the duration of the experiment. This radiation is an expected outcome from previous studies[5].

When the source SM population and a derived WS population (small-WS) were introduced sequentially, different immigration histories resulted in striking differences in the extent of diversification (Fig. 2). When the source SM was introduced first, followed at least 24 h later by small-WS, both populations increased in density (Fig. 2g, h, m, n), but diversification from the ancestral SM (Fig. 2a, b) was not significantly different from that observed in the SM-only control (Fig. 1a). In contrast, little diversification was observed when

[1]Department of Zoology, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA. [2]School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand.

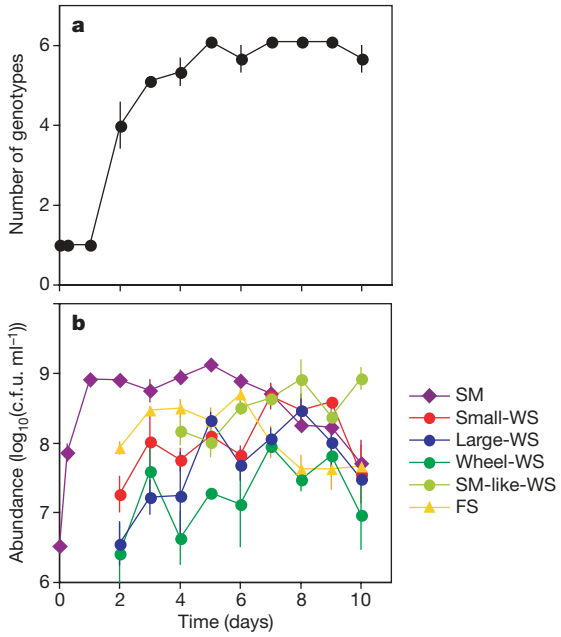


Figure 1 | Diversification of the ancestral SM genotype in spatially structured microcosms. **a**, Diversification; **b**, underlying population dynamics. Values shown are mean ± s.e.m. ($n = 3$); c.f.u., colony-forming units. See text for details of genotypes in **b**.

small-WS was introduced 6 h after (Fig. 2c), or before (Fig. 2d–f), introduction of the source SM population. Although both founding populations increased in density, diversification from the source SM was severely suppressed (Fig. 2i–l), while in most cases small-WS gave rise only to SM (Fig. 2o–r). Thus, slight differences in the early history of immigration had a large effect on the onset of adaptive radiation. This effect was solely caused by immigration history and was not due to any experimentally imposed difference in environmental conditions (compare refs 22, 25).

We further examined the dynamics of diversity over the longer term. To this end, each microcosm was homogenized after 10 days and a sample transferred to a fresh broth-filled vial. This process was repeated every 5 days (see Methods). Communities that were initially diverse showed a gradual decline in diversity and converged on a similar final diversity regardless of immigration history (Fig. 3). Boom-and-bust (or 'overshooting'[4,26]) dynamics in adaptive radiation, where species number first builds up and then declines, have been suggested to occur in a variety of taxa and habitats[4,16,25]. The radiation that occurred when the source SM population diversified in the absence of secondary invaders (Fig. 3) closely matched these dynamics (see Supplementary Discussion 1). What is even more intriguing, however, is the temporally contingent effect of secondary immigrants: early (but not late) arrival of small-WS entirely erased boom-and-bust dynamics (Fig. 3).

Changes in either the supply of mutants or the selective forces acting on these mutants (or a combination of both) are likely mechanisms of the observed effects. Three lines of evidence indicate that changes in mutation supply are not responsible. First, SM genotypes isolated from non-diversifying microcosms on day seven were indistinguishable from the wild type in terms of their capacity to generate, by mutation, WS and FS genotypes when introduced singly to fresh media (mean number of new genotypes emerged by day six ± s.e.m.: 4.83 ± 0.31 from the genotypes isolated from non-diversifying microcosms versus 4.67 ± 0.33 from the wild type; unpaired t-test, $t = 0.33$, $P > 0.74$). This result refutes the hypothesis that early arrival of small-WS induced genetic changes in SM that compromised their ability to diversify. Second, we repeated the immigration experiment using, in place of the ancestral SM genotype, a genetically derived variant (SM with a morph selection cassette, or SM^{msc}) that is phenotypically indistinguishable from SM, but engineered such that any mutation causing WS activates expression of a gene that encodes resistance to the antibiotic kanamycin (see Methods). The expression of kanamycin resistance (concomitant with mutation to WS) means that WS genotypes can be detected in a population dominated by other genotypes, simply by exposing the population to kanamycin. When small-WS was introduced early into populations founded by SM^{msc}, diversity was suppressed (as previously observed, for example, Fig. 2c, i, o) and no WS genotypes were detected when samples were spread on agar plates. However, exposure of these populations to kanamycin revealed that all microcosms contained multiple WS genotypes that were present at frequencies ranging from 1 in 10^4 to 1 in 10^6. Third, an analysis of the ability of

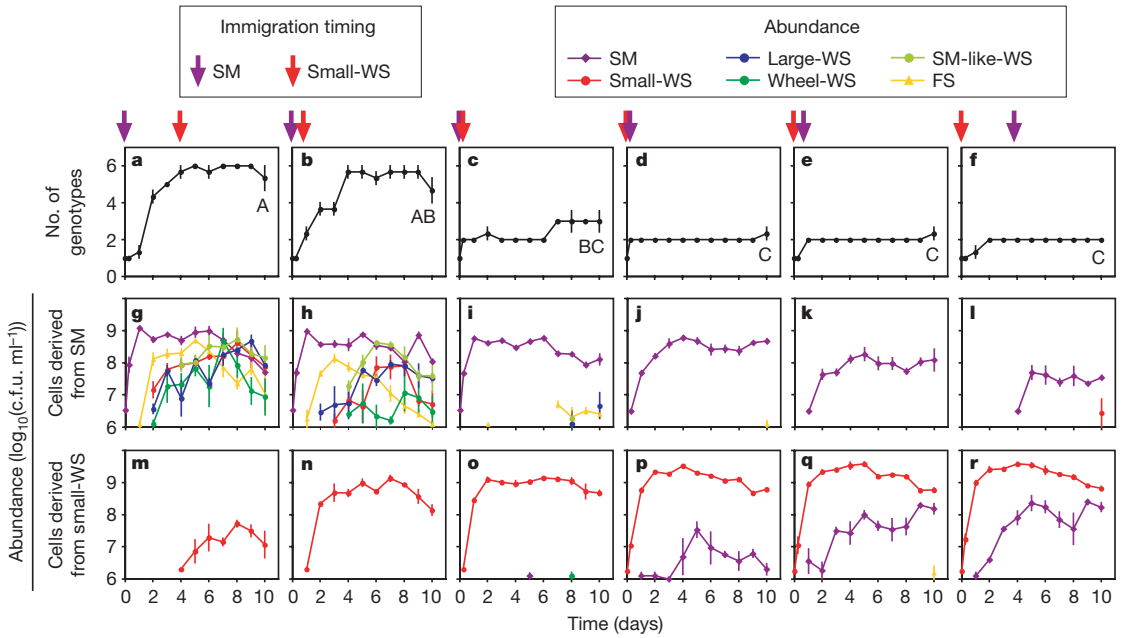


Figure 2 | Effects of immigration history. **a–f**, Effect on diversification; **g–r**, effect on population dynamics. Values shown are mean ± s.e.m. ($n = 3$). Treatments sharing the same letter (A, B, C) did not differ significantly ($P > 0.05$) in diversity on day 10 (Tukey's HSD test, $\alpha = 0.05$). Simpson and Shannon-Weaver diversity indices give similar patterns to those in **a–f**.

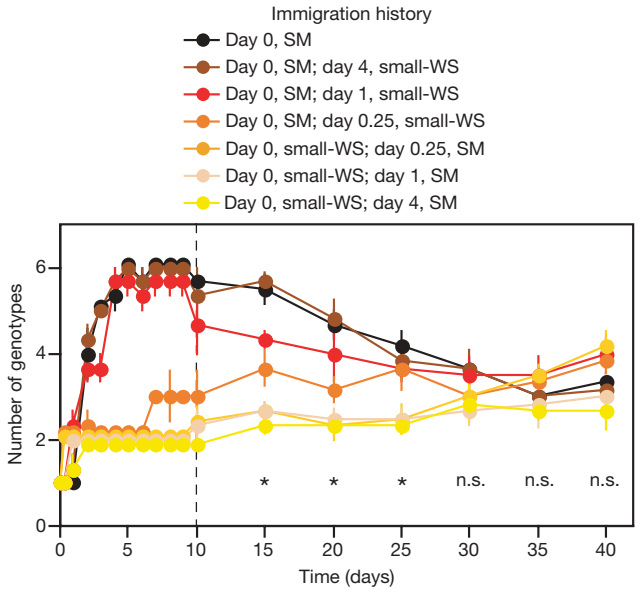


Figure 3 | Long-term changes in diversity. Dashed line indicates the start of the 5-day transfer treatments (see Methods). Immigration history had a significant effect ($P<0.05$) on diversity on days 15, 20 and 25 (asterisks), but not on days 30, 35, or 40 (n.s., not significant) (analysis of variance (ANOVA), $\alpha = 0.05$ with Bonferroni correction). Values shown are mean ± s.e.m. ($n = 3$ until day 10, and $n = 6$ after day 10).

these WS genotypes to invade populations of the ancestral SM genotype from the low frequency at which they were present in the non-diversifying microcosms showed a positive result within three days in nine out of ten instances. Together, these lines of evidence show that the suppression of diversification by early arriving immigrants is not a consequence of a marked reduction in mutation supply (see Supplementary Discussion 2).

To test the hypothesis that the introduction timing of small-WS affected the selective forces acting on the SM-derived WS genotypes, we inoculated fresh microcosms with mixtures of the four distinct WS genotypes at varying initial densities and determined the most abundant genotype after two days (see Methods). When all four genotypes were introduced at the same initial density, each became as abundant as the other three (Fig. 4). When small-WS received an advantage on account of an increased initial density, it dominated the microcosm and severely limited the growth of competing genotypes (Fig. 4). However, the other genotypes were equally capable of dominating when provided with an initial advantage (Fig. 4). These results indicate that WS genotypes are equally competitive (or neutral[3]): suppression of diversification by early arriving small-WS thus occurs as a consequence of pre-emptive colonization.

Suppression of FS by early arrival of small-WS is best explained by known competitive interactions among SM, WS and FS genotypes: FS can invade a population of SM or a mixed population of SM and WS, but not a population dominated by WS (refs 5, 23). Owing to its effect on oxygen and nutrient availability, WS has a competitive superiority over FS, such that it can exclude FS, but only in the absence of appreciable numbers of SM. In the presence of SM, WS is held in check on account of the negative frequency-dependent interaction between SM and WS, thus indirectly facilitating FS to increase in frequency[5,23]. This explains why the emergence of FS was suppressed only when small-WS was introduced early.

These mechanisms may be sufficiently general to also explain puzzling diversification patterns in other systems. For example, the filling of niches by species is thought to be deterministic in natural communities such that the same set of ecomorphs arise in multiple localities through immigration and diversification[4,14,16]. However, this expectation is not always met, even in well-described cases such as lizards on the Caribbean Islands[14], cichlids in African lakes[27], and land snails on the Hawaiian Islands[28]. In these cases, the extent of niche filling may depend stochastically on immigration history owing to indirect population interactions, as shown with FS in our study. Further, the extent of within-niche diversification is often variable among otherwise comparable communities of plants and arthropods on the Hawaiian[12], Canary[12,20] and other islands[16], which may be explained by the historically sensitive outcome of within-niche competition[3], as shown with WS here. Overall, immigration history may explain why boom-and-bust dynamics are common but not always observed, even within the same region[16].

Although it is clear that the ultimate sources of biodiversity are immigration and diversification[1–4,6], it has proven difficult to explain diversity as their additive product. Our results show that the extent of diversification can be altered greatly by even subtle differences in the early history of immigration. Such idiosyncrasies in the combined effect of immigration and diversification present a major obstacle in understanding diversity. Nevertheless, by determining the environmental conditions[10] and levels of community organization[29] under which history matters, progress can be made.

METHODS

Strains and culturing. *P. fluorescens* SBW25 (wild type) and *P. fluorescens* SBW25 *lacZ* (X.-X.Z. and P.B.R., unpublished results) were cultured in 25-ml universal vials with loose caps containing 6 ml of standard King's medium B (KB) in a 28 °C static incubator. After 4 days, cultures were plated onto KB agar and incubated for 2 days at 28 °C. Subsequently, a SM colony from the wild-type and a colony of small-WS from *P. fluorescens lacZ* were isolated, grown overnight

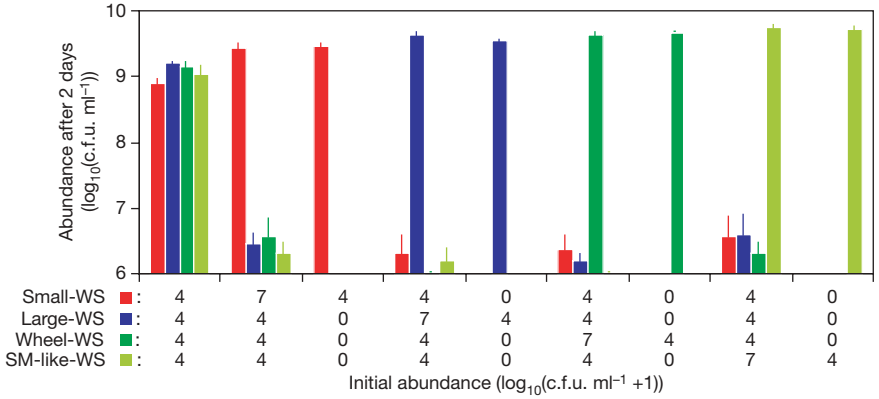


Figure 4 | Effect of the founding density of WS genotypes on the outcome of competition. Values shown are mean ± s.e.m. ($n = 3$). When the four genotypes were inoculated at the same founding density (10^4 cells ml^{-1}), they attained population sizes that were indistinguishable from one another (ANOVA, $F_{3,8} = 1.91$, $P > 0.21$). When any one of the genotypes received an advantage on account of an increased founding density (10^7 cells ml^{-1}), it became as abundant as when it was the sole colonizer (ANOVA, $F_{1,4} < 1.00$, $P > 0.37$) and severely limited the growth of the other genotypes.

in liquid KB medium at 28 °C in a 150 r.p.m. orbital shaker, and stored in 70% glycerol at −80 °C.

Manipulating immigration history. Microcosms were incubated statically with loose caps in 25-ml vials containing 6 ml KB at 28 °C. We inoculated the microcosms with SM and small-WS according to seven treatments of immigration history: (1) SM on day 0 (control), (2) SM on day 0 and small-WS on day 4, (3) SM on day 0 and small-WS on day 1, (4) SM on day 0 and small-WS on day 0.25 (that is, 6 h after SM), (5) small-WS on day 0 and SM on day 0.25, (6) small-WS on day 0 and SM on day 1, and (7) small-WS on day 0 and SM on day 4. We used 231 microcosms: 7 immigration history treatments × 3 replicates for each treatment × 11 destructive harvests. Each inoculation involved growing the stored SM and/or small WS for 16 h in liquid medium as described above and diluting so that $\sim 10^6$ cells could be introduced by transferring a volume of 20 µl into the microcosms at the pre-determined timings. Introduction of the immigrants was performed gently along the inside wall of the vials in order to minimize and standardize any resulting perturbation.

Determining genotype frequencies. Microcosms were destructively harvested 6 h after the start of the experiment and subsequently every 24 h. We determined cell densities of different genotypes by counting colonies after 2 days of growth on KB agar supplemented with 40 µg ml^{-1} 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal). We identified 6 heritable morphs: SM, small-WS, large-WS, wheel-WS, SM-like-WS and FS. Niche preference of these morphs was confirmed by observing growth of each genotype in static microcosms and establishing whether they mainly colonized the liquid phase (SM), the air–liquid interface (WS), or the bottom (FS) of the microcosms[5].

Examining long-term changes in diversity. On day 10, we transferred 6 µl of culture from each replicate of the main experiment to fresh medium. After five days, genotype frequencies were measured as described above, and 6 µl of the culture was transferred to fresh static microcosms. This procedure was repeated every 5 days until day 40.

Detecting low-frequency WS genotypes. SM^{msc} was constructed by integration of vector pUIC3 (ref. 30) harbouring the Tn*903* kanamycin resistance gene under the control of the *wss*-operon promoter (P_{wss}), into the chromosome of *P. fluorescens* SBW25 via homologous recombination. P_{wss} activity is upregulated through pleiotropic interactions in WS genotypes[24]. SM^{msc}-derived WS genotypes were detected by plating of approximately 10^6 cells from 10 independent non-diversifying microcosms onto selection medium (KB containing 30 µg ml^{-1} kanamycin, 0.0015% (w/v) Congo red and 60 µg ml^{-1} X-gal) and scored after two days. Congo red specifically stains WS colonies. Three distinct WS genotypes were identified and confirmed by re-streaking onto KB agar and by determination of the capacity of ten random WS genotypes to invade stationary-phase cultures of SM from the relative starting frequency at which they were found ($\sim 10^{-5}$). Nine out of these ten WS types were indeed capable of invading. As a control for the potential evolution of WS genotypes after plating on the selective medium, an equivalent number of cells from 6 independent 50-µl cultures (minimal selection for WS genotypes) were plated onto the selection medium. No Congo red binding colonies with WS morphologies were detected after six days of incubation, indicating that no evolution to WS occurred on the plates (see Supplementary Fig. 2).

Testing for the effect of initial densities on competitive outcome. Colonies of the four WS genotypes from day-7 plates of the SM-only inoculation treatment group were isolated and stored until use as described above. Each population was grown for 16 h in a shaken incubator, diluted as necessary, and introduced into static microcosms containing 6 ml KB at densities of 10^4 or 10^7 cells ml^{-1} (Fig. 4). Genotype frequencies were measured as described above after two days. Note that two days is long enough for populations to reach carrying capacity, but too short for evolutionary emergence of new types to confound the results.

Received 29 November 2006; accepted 29 January 2007.

1. Ricklefs, R. E. & Schluter, D. (eds) *Species Diversity in Ecological Communities: Historical and Geographical Perspectives* (Univ. Chicago Press, Chicago, 1993).
2. Losos, J. B. & Schluter, D. Analysis of an evolutionary species-area relationship. *Nature* **408,** 847–850 (2000).
3. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, 2001).
4. Gillespie, R. G. Community assembly through adaptive radiation in Hawaiian spiders. *Science* **303,** 356–359 (2004).
5. Rainey, P. B. & Travisano, M. Adaptive radiation in a heterogeneous environment. *Nature* **394,** 69–72 (1998).
6. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, Princeton, 1967).
7. Connor, E. F. & McCoy, E. D. The statistics and biology of the species-area relationship. *Am. Nat.* **113,** 791–833 (1979).
8. Drake, J. A. Community-assembly mechanics and the structure of an experimental species ensemble. *Am. Nat.* **137,** 1–26 (1991).
9. Weiher, E. & Keddy, P. A. (eds) *Ecological Assembly Rules: Perspectives, Advances, Retreats* (Cambridge Univ. Press, Cambridge, 1999).
10. Chase, J. M. Community assembly: when should history matter? *Oecologia* **136,** 489–498 (2003).
11. Fukami, T. & Morin, P. J. Productivity-biodiversity relationships depend on the history of community assembly. *Nature* **424,** 423–426 (2003).
12. Emerson, B. C. & Kolm, N. Species diversity can drive speciation. *Nature* **434,** 1015–1017 (2005).
13. Travisano, M., Mongold, J. A., Bennett, A. F. & Lenski, R. E. Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* **267,** 87–90 (1995).
14. Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K. & Rodriguez-Schettino, L. Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279,** 2115–2118 (1998).
15. Taylor, E. B. & McPhail, J. D. Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus. Proc. R. Soc. B* **267,** 2375–2384 (2000).
16. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, Oxford, 2000).
17. Gillespie, R. G., Palumbi, S. R. & Croom, H. B. Multiple origins of a spider radiation in Hawaii. *Proc. Natl Acad. Sci. USA* **91,** 2290–2294 (1994).
18. Emerson, B. C. Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. *Mol. Ecol.* **11,** 951–966 (2002).
19. Webb, C. O., Ackerly, D. D., McPeek, M. A. & Donoghue, M. J. Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* **33,** 475–505 (2002).
20. Emerson, B. C. & Oromí, P. Diversification of the forest beetle genus *Tarphius* in the Canary Islands, and the evolutionary origins of island endemics. *Evolution* **59,** 586–598 (2005).
21. Losos, J. B. & Glor, R. E. Phylogenetic comparative methods and the geography of speciation. *Trends Ecol. Evol.* **18,** 220–227 (2003).
22. Kassen, R., Llewellyn, M. & Rainey, P. B. Ecological constraints on diversification in a model adaptive radiation. *Nature* **451,** 984–988 (2004).
23. Rainey, P. B. in *The Influence of Cooperative Bacteria on Animal Host Biology* (eds McFall-Ngai, M. J., Henderson, B. & Ruby, E. G.) 83–100 (Cambridge Univ. Press, Cambridge, 2005).
24. Goymer, P. *et al.* Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. II. Role of the GGDEF regulator WspR in evolution and development of the wrinkly spreader phenotype. *Genetics* **173,** 515–526 (2006).
25. Doebeli, M. & Dieckmann, U. Speciation along environmental gradients. *Nature* **421,** 259–264 (2003).
26. Gavrilets, S. & Vose, A. Dynamic patterns of adaptive radiation. *Proc. Natl Acad. Sci. USA* **102,** 18040–18045 (2005).
27. Seehausen, O. African cichlid fish: a model system in adaptive radiation research. *Proc. R. Soc. Lond. B* **273,** 1987–1998 (2006).
28. Cowie, R. H. Variation in species diversity and shell shape in Hawaiian land snails: in situ speciation and ecological relationships. *Evolution* **49,** 1191–1202 (1995).
29. Fukami, T., Bezemer, T. M., Mortimer, S. R. & Van der Putten, W. H. Species divergence and trait convergence in experimental plant community assembly. *Ecol. Lett.* **8,** 1283–1290 (2005).
30. Rainey, P. B. Adaptation of *Pseudomonas fluorescens* to the plant rhizosphere. *Environ. Microbiol.* **1,** 243–257 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Meintjes for assistance, and B. Emerson, R. Kassen and the members of the Rainey laboratory for comments. This work was supported by the Marsden Fund Council from government funding administered by the Royal Society of New Zealand, and by the Japan Society for the Promotion of Science.

Author Contributions T.F. developed the concepts, designed the main experiment with P.B.R., collected and analysed the primary data, and wrote the manuscript in conjunction with P.B.R. and H.J.E.B. H.J.E.B. and P.B.R. conceptualized the SM^{msc} genotype, which was constructed and validated by H.J.E.B. X.-X.Z. designed, constructed and validated *lacZ*-marked SBW25. H.J.E.B. and T.F. performed selection experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.F. (tfukami@hawaii.edu).

LETTERS

Relating ligand binding to activation gating in CNGA2 channels

Christoph Biskup[1]*, Jana Kusch[1]*, Eckhard Schulz[2], Vasilica Nache[1], Frank Schwede[3], Frank Lehmann[4], Volker Hagen[5] & Klaus Benndorf[1]

Cyclic nucleotide-gated (CNG) ion channels mediate sensory signal transduction in photoreceptors and olfactory cells. Structurally, CNG channels are heterotetramers composed of either two or three homologue subunits[1–4]. Although it is well established that activation is a cooperative process of these subunits[5], it remains unknown whether the cooperativity is generated by the ligand binding, the gating, or both, and how the subunits interact. In this study, the action of homotetrameric olfactory-type CNGA2 channels[6] was studied in inside-out membrane patches by simultaneously determining channel activation and ligand binding, using the fluorescent cGMP analogue 8-DY547-cGMP as the ligand. At concentrations of 8-DY547-cGMP $< 1\,\mu M$, steady-state binding was larger than steady-state activation, whereas at higher concentrations it was smaller, generating a crossover of the steady-state relationships. Global analysis of these relationships together with multiple activation time courses following cGMP jumps[7] showed that four ligands bind to the channels and that there is significant interaction between the binding sites. Among the binding steps, the second is most critical for channel opening: its association constant is three orders of magnitude smaller than the others and it triggers a switch from a mostly closed to a maximally open state. These results contribute to unravelling the role of the subunits in the cooperative mechanism of CNGA2 channel activation and could be of general relevance for the action of other ion channels and receptors.

CNG channels are activated by the binding of cyclic nucleotides to an intracellular binding domain located in the carboxy terminus of each subunit[8]. Although insight into the function of CNG channels[9–11] and into the structure of closely related channels[12] is increasing considerably, the molecular processes underlying the translation of ligand binding to channel gating are only poorly understood. One major limitation arises from the fact that binding and gating interdepend on each other[13,14], which is a consequence of the principle of reciprocity[15,16], but no assay is known to monitor binding and gating in parallel.

For CNG channel activation all observed Hill coefficients exceed 1, ranging from 1.5 to 3.5, which means the activation gating must be a cooperative process[5]. Unfortunately, a Hill coefficient determined from electrophysiological measurements has no physical meaning and provides only a lower estimate for the number of ligands involved. The reason for this is that the Hill coefficient depends not only on binding but also on gating[13,17].

To describe activation of CNG channels more adequately and to account for the observation that these channels are already open in the absence of any ligand[18,19], investigators have used cyclic allosteric models, including the Monod–Wyman–Changeux (MWC) model[9,20] and the coupled-dimer model[10], which both assume that the binding sites of the four subunits are equivalent. In contrast, single-channel analysis of homotetrameric CNGA2 channels revealed that the binding sites are not equivalent but suggested cooperative binding of either two or three ligands[21]. Accordingly, activation time courses of CNGA2 currents, evoked by cyclic nucleotide jumps, could be described by a model with only three binding steps and pronounced positive and negative cooperativity[7]. Nevertheless, it is still a matter of debate whether the cooperativity is generated by the binding, the gating, or both, and what role the fourth subunit plays.

We studied ligand binding and activation of homotetrameric CNGA2 channels. The channels were expressed in *Xenopus laevis* oocytes, and the measurements were performed in inside-out patches of the plasma membrane, which allowed us to control the ligand concentration at the intracellular binding site. Ligand binding was evaluated under steady-state conditions by means of patch-clamp fluorometry[22,23] using a novel fluorescent cGMP derivative, 8-DY547-cGMP. The compound was synthesized by coupling the dye DY547 to position 8 of cGMP via an aminoethyl spacer (Supplementary Fig. 1a; Supplementary Methods). Functionally, the fluorophore in 8-DY547-cGMP left the capability of the cyclic nucleotide moiety to activate CNGA2 channels nearly unchanged: the concentration–activation curve for 8-DY547-cGMP was only slightly shifted to higher concentrations compared with that for cGMP (Supplementary Fig. 1e). The two ligands, 8-DY547-cGMP and cGMP, opened the channel with equal efficiency (Supplementary Fig. 2) and control of the channel activity was rapid and reversible (Supplementary Fig. 3). This allowed us to simultaneously read out channel activation and the degree of ligand binding with a freely diffusible ligand. The binding of 8-DY547-cGMP was quantified in the dome of inside-out patches using a confocal microscope (Fig. 1a–f; Supplementary Methods). The fluorescence ΔF, originating from the bound fraction of the ligand, was identified by means of a second dye, DY647, which did not bind to the channels but labelled the bath solution.

ΔF and the current amplitude, I, both measured at $+10$ mV and under steady-state conditions, were normalized with respect to the values at a saturating concentration of 20 μM (ΔF_{sat}, I_{sat}). The normalized currents (I/I_{sat}) were corrected for the small shift in the concentration–activation relationship with 8-DY547-cGMP (Supplementary Fig. 1e). The open probability, P_o, was obtained by relating I/I_{sat} to 0.99, the P_o value at saturating cGMP[7]. The result is that at [8-DY547-cGMP] $\leq 1\,\mu M$ binding exceeds activation, whereas at higher concentrations of 8-DY547-cGMP, activation exceeds binding, resulting in a crossover of the relationships (Fig. 2a). In particular, binding being lower relative to activation at [8-DY547-cGMP] $> 1\,\mu M$ is noteworthy because it directly shows that the channel is already maximally activated at submaximal liganding, suggesting

[1]Institut für Physiologie II, Friedrich-Schiller-Universität Jena, D 07740 Jena, Germany. [2]Fachhochschule Schmalkalden, Fachbereich Elektrotechnik, Blechhammer, D 98574 Schmalkalden, Germany. [3]BIOLOG Life Science Institute, Flughafendamm 9A, D 28199 Bremen, Germany. [4]Dyomics GmbH, Winzerlaer Str. 2, D 07745 Jena, Germany. [5]Leibniz-Institut für Molekulare Pharmakologie, Robert-Rössle-Str. 10, D 13125 Berlin, Germany.
*These authors contributed equally to this work.

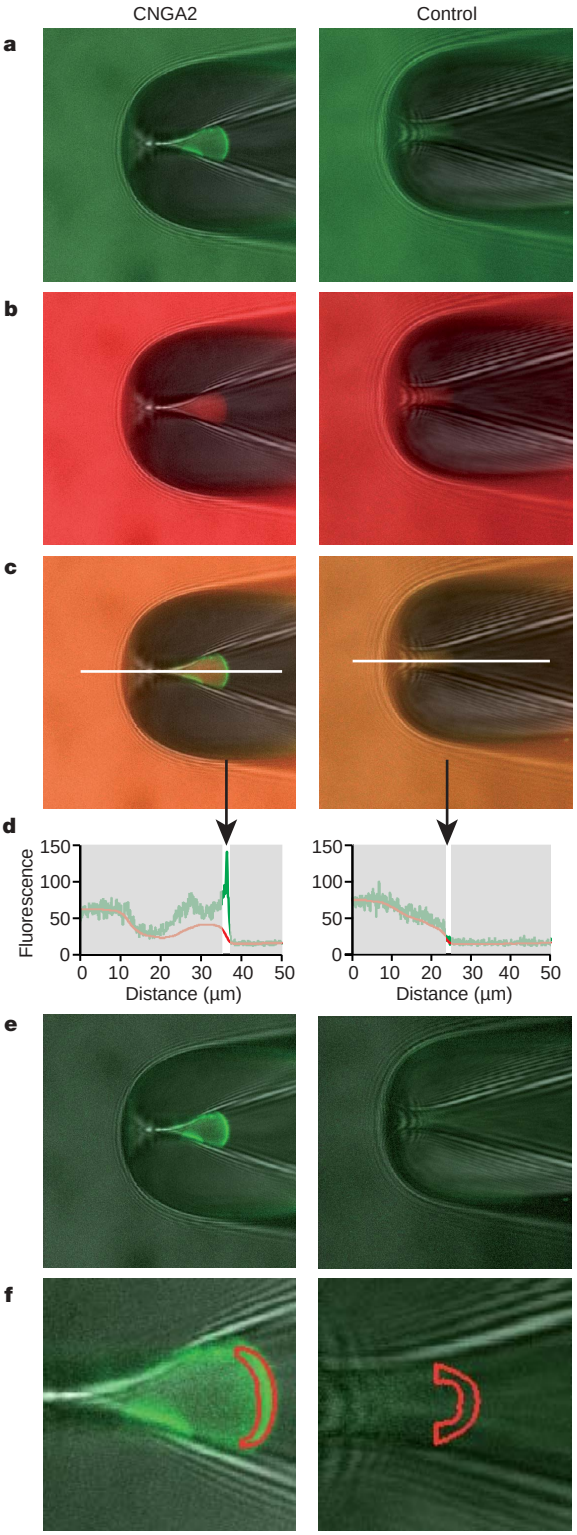


Figure 1 | Specific binding of 8-DY547-cGMP to CNGA2 channels. Confocal images of a patch containing CNGA2 channels (left panels) and a control patch from an uninjected oocyte (right panels). The bath solution contained 10 μM 8-DY547-cGMP and 1 μM DY647. **a**, Overlay of the green (8-DY547-cGMP) and the transmission channel. **b**, Overlay of the red (DY647) and the transmission channel. **c**, Overlay of the green, the red and the transmission channel. **d**, One-dimensional illustration of the scaling procedure used to quantify the portion of bound 8-DY547-cGMP along the white line in **c**. **e**, Difference image between 8-DY547-cGMP fluorescence and scaled DY647 fluorescence. **f**, Selection of the patch membrane used for analysis (red mask) by means of an automated procedure.

that a tetrameric channel is already maximally activated with only three, or possibly only two, ligands bound.

To gain further information on the gating, we studied at the same voltage of +10 mV activation time courses following cGMP concentration jumps, which were elicited by flash photolysis of DEACM-caged cGMP (see Supplementary Methods)[7]. According to previous results on CNGA2 channels[7], the activation time courses were exponential, but not sigmoidal, and they were maximally slow near the effector concentration for half-maximum response (EC_{50} value; Fig. 2b). We then globally fitted the concentration–binding relationship, the concentration–activation relationship and activation time courses at five cGMP concentrations with kinetic models by weighting each entity equally (see Supplementary Methods). We first tested the established MWC model[9]. This model assumes four identical binding sites whose affinity remains the same as long as the channel stays in the closed conformation and increases when the channel opens. The MWC model was clearly unsatisfactory (Supplementary Fig. 4a) because it not only predicts sigmoidal instead of exponential activation time courses[7], but also a higher degree of binding than activation at all ligand concentrations. A similar negative result was obtained with the coupled-dimer model.

We then examined sequential models. After discarding all models with two and three binding sites without interaction, because of unsatisfactory fits, we tested a three-state model with interaction of the binding sites in the closed state (C3L model; Supplementary Fig. 4b). This model has been shown previously to describe activation time courses reasonably[7]. Herein the C3L model was superior to the MWC and coupled-dimer model, but it did not produce the observed crossover of the relationships when globally fitted with the activation time courses. The model was therefore expanded. The most plausible expansion was to add a fourth binding step and open–closed transition, resulting in the C4L model (Fig. 3; Scheme 1 in Supplementary Methods). Like the C3L model, the C4L model produced an association constant of the second binding reaction, K_{A2}, which was by orders of magnitude smaller than the other association constants, K_{A1}, K_{A3} and K_{A4}. We could even simplify the model by setting

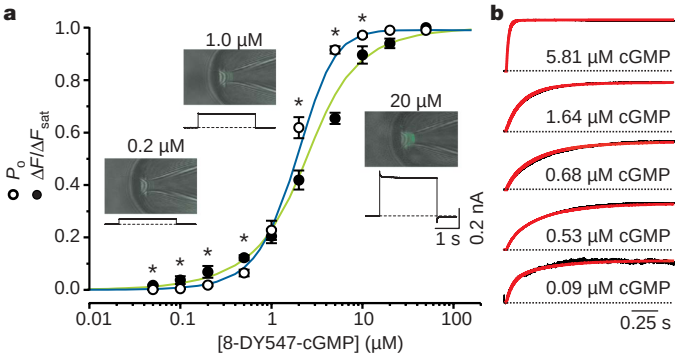


Figure 2 | Ligand binding and activation in CNGA2 channels. a, Plot of the specific binding of 8-DY547-cGMP ($\Delta F/\Delta F_{sat}$; filled circles) and activation (P_o; open circles) as a function of the 8-DY547-cGMP concentration. The data were normalized to the respective values at 50 μM 8-DY547-cGMP. Each data point is the mean of measurements obtained from 6 to 15 patches. Error bars indicate s.e.m. Asterisks indicate significant ($P < 0.05$) differences between binding and activation: binding exceeds activation at [8-DY547-cGMP] < 1 μM and activation exceeds binding at [8-DY547-cGMP] > 1 μM. The inserts show examples at three 8-DY547-cGMP concentrations of simultaneously recorded current traces and confocal difference images, reflecting the bound 8-DY547-cGMP. All recordings were obtained from the same patch. The voltage was stepped from 0 to +10 mV. The curves represent the best global fit with the C4L model (green, binding; blue, activation). **b**, Time course of activation following cGMP jumps from zero to the indicated cGMP concentrations. The black traces are averages of normalized traces obtained from four different patches. The red curves represent the best global fit with the C4L model (Fig. 3; Scheme 1 in Supplementary Methods).

$K_{A1} = K_{A3} = K_{A4}$ and still obtained a significantly better fit with the C4L than the C3L model though the number of parameters was equal. In particular, the C4L model produced the crossover of the steady-state relationships. All equilibrium association constants, equilibrium constants for closed–open isomerizations, and rate constants for the closed–open transitions were well determined (see Supplementary Methods). The rate constants of binding and unbinding are apparently too fast to impose any constraints on the gating and could thus not be determined by our analysis.

These results lead to the following conclusions: (1) the closed–open transitions are solely rate limiting for the activation time courses; (2) the binding of the first ligand generates a small but noticeable P_o of 0.04; (3) the equilibrium association constant for the second ligand (K_{A2}) is about three orders of magnitude smaller than the other three equilibrium association constants, indicating pronounced negative and positive cooperativity for the ligand binding; and (4) the binding of the second ligand causes the main opening switch to a value close to 1. In other words, the gating is a remarkably simple process: the first ligand induces small but noticeable opening, the second ligand switches the still mostly closed to a fully open channel, whereas the third and fourth ligand only drive the channel away from the critical second binding step, stabilizing the open conformation (Fig. 3).

What could be the physical mechanism of binding sites interacting in a closed channel? It is intriguing to assume that the cyclic-nucleotide-binding domain of the subunit binding the first ligand exerts a conformational change, which spreads to other subunits, thereby changing their binding affinity. The association constants K_{Ax} in our C4L model would then characterize the respective equilibria of ligand binding and unbinding plus the respective conformational change. It remains unresolved here whether after the binding of the first ligand either one of the two adjacent subunits or the opposite subunit changes its affinity (Fig. 3). Previous data suggest that liganding of one subunit modulates only the binding affinity of adjacent subunits[24].

Our results resemble to some extent the results of some researchers[10], who show for closely related CNGA1 channels that binding of a single ligand causes noticeable opening ($P_o = 0.017$), but differ from those of other researchers[25,26], who suggest that single liganding does not open CNGA1 channels. In contrast to our results on CNGA2 channels, both studies conclude that four ligands are required for full channel activation. One has to take into account, however, that CNGA2 channels are only similar, but not equal, to CNGA1 channels and, in addition, that both groups worked with experimentally altered CNGA1 channels, either by covalently tethering the ligands[25,26] or by extensive mutagenesis[10]. It is therefore more expected than surprising that these results differ from ours and also from each other.

A striking similarity exists between activation gating in CNGA2 channels and the glycine receptor: The natural glycine receptor is a pentamer, consisting of three α plus two β subunits[27], of which only the three α subunits are required to describe activation gating in homopentameric channels[28,29]. In analogy, the natural olfactory channel is a tetramer, consisting of two CNGA2 subunits plus the two regulatory subunits CNGA4 and CNGB1b[30], of which two CNGA2 subunits are required to describe activation gating in homotetrameric channels. This is particularly appealing not only because of this similarity between a pentamer and a tetramer, but also because the molecular structures of the subunits are unrelated. One can therefore speculate that in natural heteromeric channels only those subunits that can form homomeric channels do the basic gating work. Despite these similarities to the glycine receptors, our results differ in one respect substantially: the equilibrium constants of association are neither independent nor do they change monotonically but strongly depend on each other.

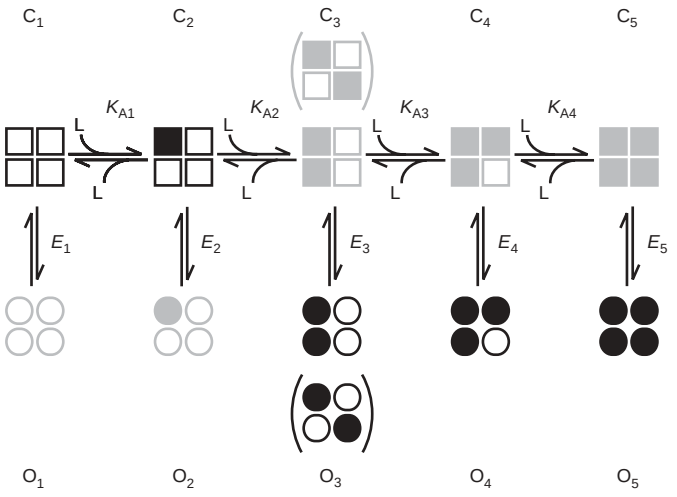


Figure 3 | Mechanistic scheme of the C4L model. K_{Ax} and E_x are the equilibrium constants for ligand association and closed–open isomerization; squares and circles represent subunits in a closed (C_x) or open (O_x) channel; open and filled symbols represent unliganded and liganded subunits, respectively. L is a ligand molecule. The shade of grey indicates in which direction the equilibrium of each closed–open isomerization is shifted. E_1 and E_5 were determined from single-channel recordings[7] to be 1.20×10^{-4} and $E_5 = 9.90 \times 10^{1}$, respectively. The equilibrium constants determined by the global fit are $K_{A1} = K_{A3} = K_{A4} = (1.01 \pm 0.04) \times 10^{6}\,M^{-1}$; $K_{A2} = (2.94 \pm 0.15) \times 10^{2}\,M^{-1}$; $E_2 = (3.89 \pm 0.50) \times 10^{-2}$; $E_3 = (1.17 \pm 0.09) \times 10^{3}$; $E_4 = (5.68 \pm 0.63) \times 10^{2}$ and reduced $\chi^2 = 2.32$.

METHODS

The specific binding of 8-DY547-cGMP to inside-out patches was performed using a confocal microscope (Fig. 1; Supplementary Methods). The cytosolic face of the patch was perfused with a solution containing the desired concentration of 8-DY547-cGMP and the red dye DY647 as reference to determine the spatial distribution of unbound ligand molecules.

In patches containing CNGA2 channels, the fluorescence of bound 8-DY547-cGMP exceeded the fluorescence of the surrounding bath solution, delineating the patch. Control patches did not show an enhanced fluorescence signal (Fig. 1a). In the red channel, enhanced fluorescence was neither observed in the presence nor absence of CNGA2 channels (Fig. 1b). Overlay of both channels yielded orange colour except for the green regions where 8-DY547-cGMP fluorescence exceeded DY647 fluorescence, owing to binding to CNGA2 channels (Fig. 1c). The procedure for determining the intensity of bound 8-DY547-cGMP fluorescence is illustrated in Fig. 1d for the one-dimensional case (white line in Fig. 1c). Offset and scaling of the DY647 fluorescence were adjusted such that DY647 intensities matched 8-DY547-cGMP intensities in the interior of the pipette (where no dye is present) and in the bath solution (where both dyes are equally distributed). The difference between the green and red profile is caused by specific binding of 8-DY547-cGMP to the channels. To visualize the fraction of bound 8-DY547-cGMP in the whole image, the difference (ΔF) between the measured green and the scaled red profile was determined for each pixel (Fig. 1e). Accordingly, ΔF was near zero in the bath solution and also in control patches, whereas bright signals originated from patches containing CNGA2 channels. To determine the 8-DY547-cGMP concentration–binding curve, average intensities in regions belonging to the electrophysiologically relevant dome of the patch were determined by generating a mask with a software algorithm that identified pixels for which the DY647 fluorescence gradients were steepest.

The technique used for flash photolysis[7] and all other methods are described in Supplementary Methods.

Received 24 October 2006; accepted 12 January 2007.
Published online 25 February 2007.

1. Körschen, H. G. *et al.* A 240 kDa protein represents the complete β subunit of the cyclic nucleotide-gated channel from rod photoreceptor. *Neuron* **15**, 627–636 (1995).
2. Bönigk, W. *et al.* The native rat olfactory cyclic nucleotide-gated channel is composed of three distinct subunits. *J. Neurosci.* **19**, 5332–5347 (1999).
3. Zhong, H., Molday, L. L., Molday, R. S. & Yau, K.-W. The heteromeric cyclic nucleotide-gated channel adopts a 3A:1B stoichiometry. *Nature* **420**, 193–198 (2002).
4. Zheng, J. & Zagotta, W. N. Stoichiometry and assembly of olfactory cyclic nucleotide-gated channels. *Neuron* **42**, 411–421 (2004).

5. Kaupp, U. B. & Seifert, R. Cyclic nucleotide-gated ion channels. *Physiol. Rev.* **82,** 769–824 (2002).
6. Dhallan, R. S., Yau, K.-W., Schrader, K. A. & Reed, R. R. Primary structure and functional expression of a cyclic nucleotide-activated channel from olfactory neurons. *Nature* **347,** 184–187 (1990).
7. Nache, V. *et al.* Activation of olfactory-type cyclic nucleotide-gated channels is highly cooperative. *J. Physiol. (Lond.)* **569,** 91–102 (2005).
8. Kaupp, U. B. *et al.* Primary structure and functional expression from complementary DNA of the rod photoreceptor cyclic GMP-gated channel. *Nature* **342,** 762–766 (1989).
9. Goulding, E. H., Tibbs, G. R. & Siegelbaum, S. A. Molecular mechanism of cyclic-nucleotide-gated channel activation. *Nature* **372,** 369–374 (1994).
10. Liu, D. T., Tibbs, G. R., Paoletti, P. & Siegelbaum, S. A. Constraining ligand-binding site stoichiometry suggests that a cyclic nucleotide-gated channel is composed of two functional dimers. *Neuron* **21,** 235–248 (1998).
11. Liu, D. T. & Siegelbaum, S. A. Change of pore helix conformational state upon opening of cyclic nucleotide-gated channels. *Neuron* **28,** 899–909 (2000).
12. Zagotta, W. N. *et al.* Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature* **425,** 200–205 (2003).
13. Colquhoun, D. Binding, gating, affinity and efficacy: the interpretation of structure–activity relationships for agonists and of the effects of mutating receptors. *Br. J. Pharmacol.* **125,** 924–947 (1998).
14. Colquhoun, D. Agonist-activated ion channels. *Br. J. Pharmacol.* **147,** S17–S26 (2006).
15. Edsall, J. T. & Wyman, J. *Biophysical Chemistry*. (Academic Press, New York, 1958).
16. Wyman, J. & Gill, S. J. *Binding and Linkage. Functional Chemistry of Biological Macromolecules.* (Univ. Science Books, Mill Valley, 1990).
17. Li, J., Zagotta, W. N. & Lester, H. A. Cyclic-nucleotide gated channels: structural basis of ligand efficacy and allosteric modulation. *Q. Rev. Biophys.* **30,** 177–193 (1997).
18. Picones, A. & Korenbrot, J. I. Spontaneous ligand-independent activity of the cGMP-gated ion channels in cone photoreceptors in fish. *J. Physiol. (Lond.)* **485,** 699–714 (1995).
19. Tibbs, G. R., Goulding, E. H. & Siegelbaum, S. A. Allosteric activation and tuning of ligand efficacy in cyclic-nucleotide-gated channels. *Nature* **386,** 612–615 (1997).
20. Monod, J., Wyman, J. & Changeux, J. P. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12,** 88–118 (1965).
21. Li, J. & Lester, H. A. Single-channel kinetics of the rat olfactory cyclic nucleotide-gated channel expressed in *Xenopus* oocytes. *Mol. Pharmacol.* **55,** 883–893 (1999).
22. Zheng, J. & Zagotta, W. N. Patch-clamp fluorometry recording of conformational rearrangements of ion channels. *Sci. STKE* 10.1126/stke.2003.176.pl7 (2003).
23. Trudeau, M. C. & Zagotta, W. N. Dynamics of Ca^{2+}-calmodulin-dependent inhibition of rod cyclic nucleotide-gated channels measured by patch-clamp fluorometry. *J. Gen. Physiol.* **124,** 211–223 (2004).
24. Craven, K. B. & Zagotta, W. N. Salt bridges and gating in the COOH-terminal region of HCN2 and CNGA1 channels. *J. Gen. Physiol.* **124,** 663–677 (2004).
25. Ruiz, M. L. & Karpen, J. W. Single cyclic nucleotide-gated channels locked in different ligand-bound states. *Nature* **389,** 389–392 (1997).
26. Ruiz, M. L. & Karpen, J. W. Opening mechanism of a cyclic nucleotide-gated channel based on analysis of single channels locked in each liganded state. *J. Gen. Physiol.* **113,** 873–895 (1999).
27. Brejc, K. *et al.* Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411,** 269–276 (2001).
28. Beato, M., Groot-Kormelink, P. J., Colquhoun, D. & Sivilotti, L. G. The activation mechanism of α1 homomeric glycine receptors. *J. Neurosci.* **24,** 895–906 (2004).
29. Burzomato, V., Beato, M., Groot-Kormelink, P. J., Colquhoun, D. & Sivilotti, L. G. Single-channel behaviour of heteromeric α1β glycine receptors: An attempt to detect a conformational change before the channel opens. *J. Neurosci.* **24,** 10924–10940 (2004).
30. Bradley, J., Bönigk, W., Yao, K.-W. & Frings, S. Calmodulin permanently associates with rat olfactory CNG channels under native conditions. *Nature Neurosci.* **7,** 705–710 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are indebted to U. B. Kaupp, Forschungszentrum Jülich, for providing the CNGA2 clone. We also thank A. Hertel, K. Schoknecht, S. Bernhardt, A. Kolchmeier and B. Tietsch for technical assistance. This work was supported by grants of the Deutsche Forschungsgemeinschaft to K.B. F.S. acknowledges support of BIA, Bremen.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.B. (klaus.benndorf@mti.uni-jena.de).

LETTERS

p53-induced inhibition of Hif-1 causes cardiac dysfunction during pressure overload

Masanori Sano[1]*, Tohru Minamino[1]*, Haruhiro Toko[1], Hideyuki Miyauchi[1], Masayuki Orimo[1], Yingjie Qin[1], Hiroshi Akazawa[1], Kaoru Tateno[1], Yosuke Kayama[1], Mutsuo Harada[1], Ippei Shimizu[1], Takayuki Asahara[2], Hirofumi Hamada[3], Shuhei Tomita[4], Jeffrey D. Molkentin[5], Yunzeng Zou[6] & Issei Komuro[1]

Cardiac hypertrophy occurs as an adaptive response to increased workload to maintain cardiac function[1]. However, prolonged cardiac hypertrophy causes heart failure[2], and its mechanisms are largely unknown. Here we show that cardiac angiogenesis is crucially involved in the adaptive mechanism of cardiac hypertrophy and that p53 accumulation is essential for the transition from cardiac hypertrophy to heart failure. Pressure overload initially promoted vascular growth in the heart by hypoxia-inducible factor-1 (Hif-1)-dependent induction of angiogenic factors, and inhibition of angiogenesis prevented the development of cardiac hypertrophy and induced systolic dysfunction. Sustained pressure overload induced an accumulation of p53 that inhibited Hif-1 activity and thereby impaired cardiac angiogenesis and systolic function. Conversely, promoting cardiac angiogenesis by introducing angiogenic factors or by inhibiting p53 accumulation developed hypertrophy further and restored cardiac dysfunction under chronic pressure overload. These results indicate that the anti-angiogenic property of p53 may have a crucial function in the transition from cardiac hypertrophy to heart failure.

During the development of cardiac hypertrophy, it has been postulated that a mismatch between the number of capillaries and the size of cardiomyocytes develops, leading to myocardial hypoxia[3,4]. There are various reports indicating a potential relationship between cardiac angiogenesis, cardiac hypertrophy and cardiac function[5–7]. We thus proposed that cardiac angiogenesis might contribute to the development of cardiac hypertrophy and that its impairment might induce heart failure.

We first established a hypertrophy model that shows cardiac dysfunction at chronic stage by performing a severe transverse aorta constriction (TAC). In this model, cardiac hypertrophy gradually developed, reached a peak on day 14 after TAC and decreased afterwards (Fig. 1a–d). Fractional shortening was preserved until day 14 and significantly decreased on day 28 with left ventricular dilation and increased cardiac fibrosis (Fig. 1d, e, and Supplementary Fig. 1a, b). These results suggest that pressure overload initially induced 'adaptive' hypertrophy (days 1–14) with preserved cardiac function; however, this adaptive mechanism could not protect the hypertrophied heart against sustained pressure overload, resulting in systolic dysfunction (days 14–28). The number of microvessels per cardiomyocyte increased until day 14 and decreased thereafter (Fig. 1b, c). The number of bromodeoxyuridine (BrdU)-positive endothelial cells, but not that of BrdU-positive cardiomyocytes, was significantly increased (Supplementary Fig. 1c). Consistent with these results was our observation that the expression of angiogenic factors such as vascular endothelial growth factor (VEGF) and angiopoietin-1 (Ang-1) was upregulated in the early phase and decreased in the late phase (Fig. 1f, g).

To explain the role of angiogenesis in the development of cardiac hypertrophy, we examined the effects of TNP-470, an inhibitor of angiogenesis[8], on cardiac hypertrophy. TNP-470 suppressed the increase in the number of microvessels in the heart of mice that had undergone TAC (Fig. 2a). TAC-induced hypertrophy was almost completely inhibited by the treatment with TNP-470 (Fig. 2a, b). Administration of TNP-470 to mice significantly impaired cardiac function at 2 weeks after TAC (Fig. 2b). Similar inhibitory effects of TNP-470 were observed in other models of cardiac hypertrophy such as an angiotensin II infusion model (Supplementary Fig. 2a–d). These results suggest that cardiac angiogenesis is crucially involved in preserving cardiac function as well as in developing cardiac hypertrophy.

To examine whether promoting angiogenesis prevents the transition from cardiac hypertrophy towards heart failure, we introduced adenoviral vectors encoding VEGF and Ang-1 directly into the heart and produced pressure overload. Introduction of angiogenic factors enhanced an increase in the number of microvessels compared with that of LacZ after TAC (Fig. 2c). Cardiac hypertrophy was further developed and its function was preserved in the VEGF/Ang-1 group at 4 weeks after TAC (Fig. 2c, d). Conversely, the introduction of a soluble form of Flt-1, an inhibitor of angiogenesis[9], into the thigh muscle markedly reduced cardiac hypertrophy as well as the number of microvessels compared with that of LacZ, and suppression of this adaptive response caused a further decline in cardiac systolic function at 4 weeks after TAC (Fig. 2c, d). These results indicate that cardiac angiogenesis, which is induced in the early adaptive phase, may be sufficient to maintain cardiac function and that the angiogenesis becomes insufficient to keep the function of hypertrophied hearts in the maladaptive phase, presumably because of decreased expression of angiogenic factors.

Cardiomyocyte hypertrophy has been thought to increase diffusion distance, resulting in reduced oxygen supply in the myocardium. Neovascularization associated with cardiac hypertrophy may be attributable to angiogenic factors in cardiomyocytes being upregulated by hypoxia. We therefore examined the expression of Hif-1α, a key transcription factor for the hypoxic induction of angiogenic

[1]Department of Cardiovascular Science and Medicine, Chiba University Graduate School of Medicine, 1-8-1 Inohana, Chuo-ku, Chiba 260-8670, Japan. [2]Stem Cell Translational Research, Kobe Institute of Biomedical Research and Innovation/RIKEN Center for Developmental Biology, 2-2 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan. [3]Department of Molecular Medicine, Sapporo Medical University, S1 W17, Chuo-ku, Sapporo 060-8556, Japan. [4]Division of Experimental Immunology, Institute for Genome Research, University of Tokushima, 3-18-15 Kuramoto, Tokushima 770-8503, Japan. [5]Department of Pediatrics, Children's Hospital Medical Center, Division of Molecular Cardiovascular Biology, 3333 Burnet Avenue, Cincinnati, Ohio 45229-3039, USA. [6]Shanghai Institute of Cardiovascular Diseases, Zhongshan Hospital and Institutes of Biomedical Sciences, Fudan University, 180 Feng Lin Road, Shanghai 200032, China.
*These authors contributed equally to this work.

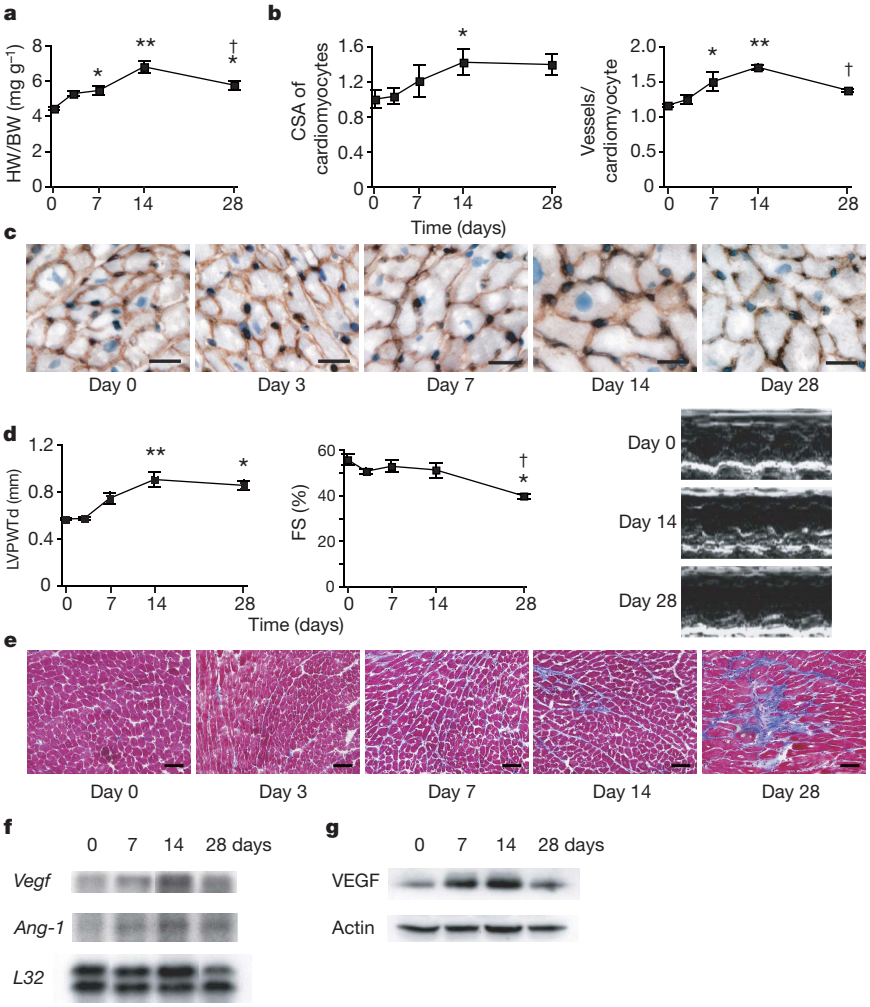


Figure 1 | Cardiac hypertrophy, function and angiogenesis after TAC. **a**, Heart weight/body weight (HW/BW) after TAC. **b**, Cross-sectional area (CSA) of cardiomyocytes and the number of microvessels per cardiomyocyte. **c**, Double-immunostaining for dystrophin (brown) and PECAM (black) of the TAC hearts. Scale bar, 20 μm. **d**, Echocardiographic analysis. FS, fractional shortening; LVPWTd, left ventricular posterior wall thickness. Asterisk, $P < 0.05$, two asterisks, $P < 0.001$ versus day 0; dagger, $P < 0.01$ versus day 14. Error bars indicate s.e.m.; $n = 7$ for **a**; $n = 3$ for **b**; $n = 5$ for **d**. **e**, Cardiac fibrosis. Scale bar, 50 μm. **f**, **g**, RNase protection assay (**f**) and western blot analysis (**g**) in the TAC heart.

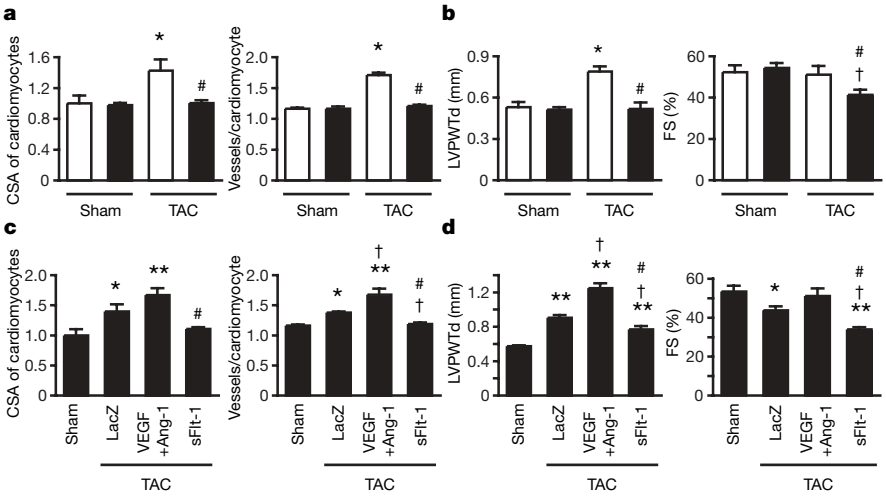


Figure 2 | Cardiac angiogenesis in TAC-induced hypertrophy. **a**, **b**, Mice were subjected to TAC or sham operation and treated with (filled columns) or without (open columns) TNP-470 for 2 weeks. Histological (**a**) and echocardiographic (**b**) analyses were performed. CSA, cross-sectional area; FS, fractional shortening; LVPWTd, left ventricular posterior wall thickness. Asterisk, $P < 0.005$ versus sham without TNP-470; dagger, $P < 0.01$ versus sham with TNP-470; hash sign, $P < 0.05$ versus TAC without TNP-470. Error bars indicate s.e.m.; $n = 3$ for **a**; $n = 5$ for **b**. **c**, **d**, Mice were infected with adenoviral vectors encoding VEGF and Ang-1, soluble Flt-1 (sFlt-1) or LacZ and subjected to 4 weeks of TAC or sham operation. Histological (**c**) and echocardiographic (**d**) analyses were performed. Asterisk, $P < 0.05$, two asterisks, $P < 0.005$ versus sham; dagger, $P < 0.05$ versus TAC with LacZ; hash sign, $P < 0.005$ versus TAC with VEGF + Ang-1. Error bars indicate s.e.m.; $n = 4$ for **c**; $n = 3$ for **d**.

factors, in the hearts of the TAC model. Expression of Hif-1α, but not that of Hif-2α, was increased by pressure overload from day 3 (Fig. 3a, and Supplementary Fig. 3a). Similarly, Hif-1 activity was increased from day 3 to day 14, indicating an important function for Hif-1α in the adaptive mechanism of cardiac hypertrophy (Fig. 3b, and Supplementary Fig. 4). To test this, we employed the conditional knockout mice in which the deletion of *Hif-1α* could be induced by tamoxifen only in cardiomyocytes (Supplementary Fig. 5a, b). Two weeks after TAC, both the expression of VEGF and the number of microvessels were significantly lower in the *Hif-1α* mutant mice than in control mice (Fig. 3c, d). Similarly, the development of cardiac hypertrophy was significantly attenuated in the *Hif-1α* mutant mice in comparison with control mice (Fig. 3d, e). Cardiac function in the *Hif-1α* mutant mice, but not in control mice, was significantly impaired at 2 weeks after TAC (Fig. 3e).

Our results indicate that the downregulation of Hif-1α and the resulting decreased angiogenic factors may cause maladaptive hypertrophy during chronic pressure overload. We next examined the extent of ischaemia in the hypertrophied myocardium with a hypoxyprobe[10]. Pressure overload induced cardiac hypoxia from as early as 3 days after TAC, and cardiac ischaemia was sustained until day 28 (Fig. 3f), indicating that a decrease in Hif-1α expression in the maladaptive phase was not due to an improvement of hypoxia in the myocardium by neoangiogenesis. It is generally accepted that Hif-1α is stabilized under mild or moderate hypoxia[11,12]. Prolonged or more severe hypoxia has been reported to induce the tumour suppressor p53, which binds to Hif-1α, promoting its degradation and inhibiting its transactivation properties[13,14]. We therefore examined whether p53 accumulates in the myocardium after chronic pressure overload, thereby inhibiting the Hif-1-dependent induction of angiogenic factors. Expression of p53 was markedly upregulated at 14 days after TAC, in accordance with an increase in atrial natriuretic factor (Anf), one of the markers for cardiac dysfunction (Fig. 3g, and Supplementary Fig. 4). Levels of phosphorylated p53 were also elevated in the maladaptive phase, whereas messenger RNA levels were unchanged (Fig. 3g, and Supplementary Fig. 3b). Chronic pressure overload increased the phosphorylation of Chk2, a protein kinase that mediates the p53-dependent DNA repair pathways (Fig. 3g). Expression of Bax, a pro-apoptotic factor regulated by p53, was also upregulated by sustained pressure overload (Fig. 3g), whereas

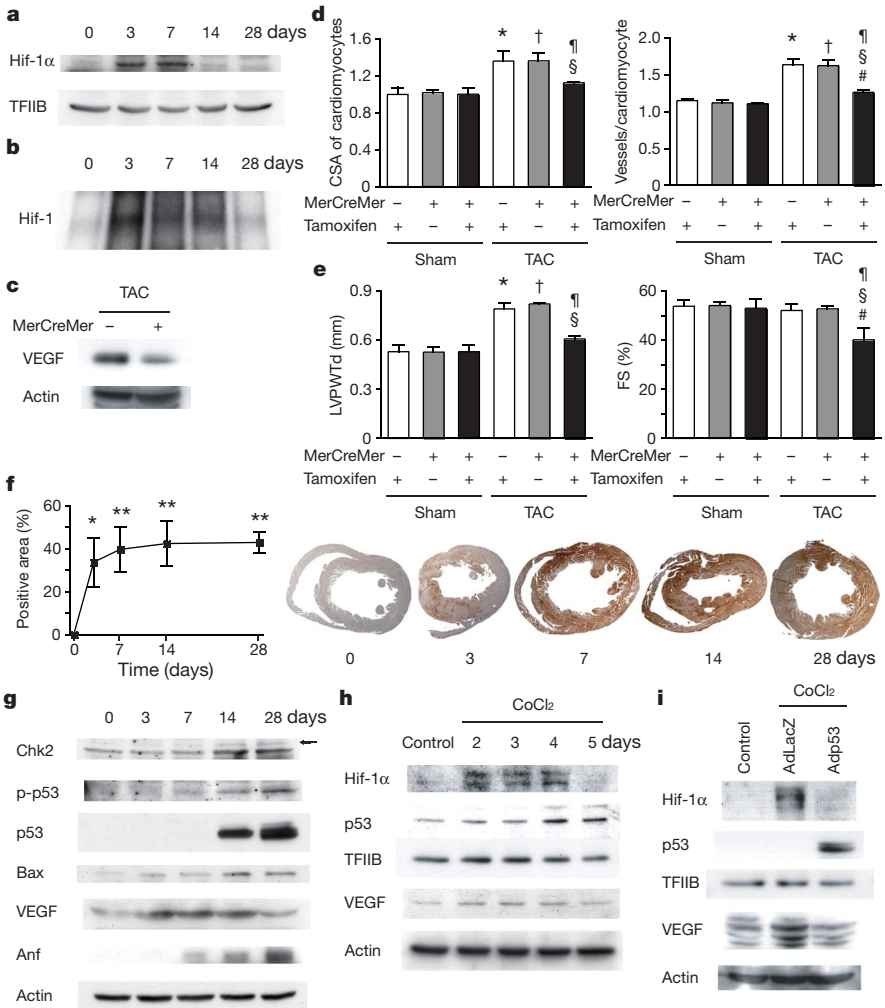


Figure 3 | Role of Hif-1 in adaptive hypertrophy. **a**, **b**, Hif-1α expression (**a**) and Hif-1 activity (**b**) in hearts during 4 weeks of TAC. **c**, Western blot analysis in the heart of mutant (+) or control (−) mice subjected to TAC and tamoxifen treatment. **d**, **e**, Histological (**d**) and echocardiographic (**e**) analyses of mutant mice (MerCreMer (+)/Tamoxifen (+)) or control littermates (MerCreMer (+)/Tamoxifen (−) or MerCreMer (−)/Tamoxifen (+)) subjected to 2 weeks of TAC or sham operation. CSA, cross-sectional area; FS, fractional shortening; LVPWTd, left ventricular posterior wall thickness. Asterisk, $P < 0.005$ versus sham/MerCreMer (−)/tamoxifen (+); dagger, $P < 0.005$ versus sham/MerCreMer (+)/tamoxifen (−); hash sign, $P < 0.05$ versus sham/MerCreMer (+)/tamoxifen (+); section sign, $P < 0.05$ versus TAC/MerCreMer (−)/tamoxifen (+); paragraph sign, $P < 0.05$ versus TAC/MerCreMer (+)/tamoxifen (−). Error bars indicate s.e.m.; $n = 4$. **f**, Cardiac ischaemia (brown). Asterisk, $P < 0.05$, two asterisks, $P < 0.01$ versus day 0. Error bars indicate s.e.m.; $n = 6$. **g**, Western blot analysis in hearts during 4 weeks of TAC. The arrow indicates the position of phospho-Chk2. **h**, Expression of Hif-1α, p53 and VEGF in primary cultured cardiomyocytes treated with $CoCl_2$. **i**, Expression of Hif-1α and p53 in cultured cardiomyocytes infected with an adenoviral vector encoding p53 (Adp53) or mock (AdLacZ) and treated with $CoCl_2$.

expression of anti-angiogenic factors such as thrombospondin-1 and plasminogen activator inhibitor-1 was unchanged (Supplementary Fig. 3c, d).

Treatment of cultured cardiomyocytes with $CoCl_2$, a hypoxia mimetic, induced the expression of Hif-1α and VEGF during days 2–4, but their expression was downregulated thereafter even in the presence of $CoCl_2$ (Fig. 3h). In contrast, p53 was increased at 4 days after treatment (Fig. 3h). Forced expression of p53 markedly inhibited the upregulation of Hif-1α and VEGF induced by $CoCl_2$ (Fig. 3i). Similar results were observed in cardiomyocytes exposed to hypoxia (Supplementary Fig. 6a, b). The introduction of p53 markedly attenuated Hif-1α expression, which was blocked by treatment with the proteasome inhibitor MG132 (Supplementary Fig. 6c). Hif-1α was co-immunoprecipitated with p53 in the presence of MG132 (Supplementary Fig. 6c). Treatment with $CoCl_2$ or hypoxia significantly upregulated the Hif-1 reporter activity (Supplementary Fig. 6d, e), and this upregulation was significantly suppressed by the introduction of p53 (Supplementary Fig. 6d, e), indicating a possibly crucial role of p53 in the downregulation of Hif-1α transcriptional activity.

To investigate the role of p53 in maladaptive hypertrophy, we produced the TAC model in p53-deficient mice. Cardiac Hif-1 activity and VEGF levels were higher in these mice at 4 weeks after TAC than in wild-type mice (Fig. 4a, b, and Supplementary Fig. 7). Consequently, the number of microvessels became significantly greater in the heart of p53-deficient mice than in wild-type mice 4 weeks after the operation (Fig. 4c). After chronic pressure overload, p53-deficient mice had more marked cardiac hypertrophy, better systolic function and lower Anf levels than wild-type mice (Fig. 4b, c, and Supplementary Fig. 7). Inhibition of p53 also has beneficial effects on cardiac function after myocardial infarction, and these effects may be partly attributable to increased neovascularization (Supplementary Fig. 8a–c).

We next used the p53 activator quinacrine[15] to examine whether p53 activation in the early phase of pressure overload promotes the transition of cardiac hypertrophy to heart failure. Administration of quinacrine significantly increased levels of p53 in the heart of sham-operated mice and enhanced the accumulation of p53 by TAC (Fig. 4d, and Supplementary Fig. 9). The induction of Hif-1 activity and VEGF expression by TAC was significantly suppressed in the heart of quinacrine-treated mice (Fig. 4d, e, and Supplementary Fig. 9). Consequently, p53 upregulated by quinacrine treatment severely impaired cardiac angiogenesis induced by TAC and attenuated adaptive hypertrophy, leading to systolic dysfunction associated with increased Anf levels at 2 weeks after the operation (Fig. 4d, f, and Supplementary Fig. 9). These effects of quinacrine were blunted in p53-deficient mice (Supplementary Fig. 10a, b). The number of cardiomyocytes positive for TdT-mediated dUTP nick end labelling

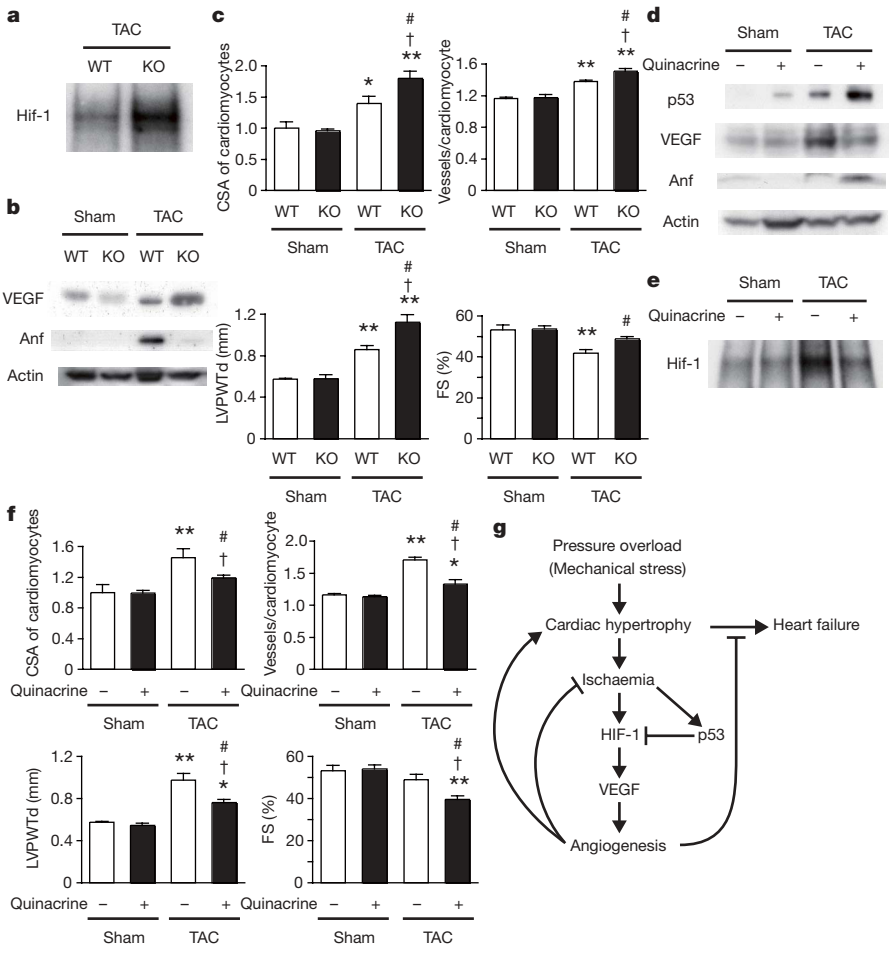


Figure 4 | Role of p53 accumulation in maladaptive hypertrophy. **a–c**, Wild-type (WT) and p53-deficient mice (KO) were subjected to 4 weeks of TAC or sham operation. Gel mobility-shift assay (**a**), western blotting (**b**) and histological and echocardiographic analyses (**c**) were performed. CSA, cross-sectional area; FS, fractional shortening; LVPWTd, left ventricular posterior wall thickness. Asterisk, $P < 0.05$, two asterisks, $P < 0.005$ versus wild-type sham; dagger, $P < 0.001$ versus KO sham; hash sign, $P < 0.05$ versus wild-type TAC. Error bars indicate s.e.m.; $n = 4$. **d–f**, Wild-type mice were treated with quinacrine (+) or vehicle (−) and subjected to 2 weeks of TAC or sham operation. Western blot analysis (**d**), gel mobility-shift assay (**e**) and histological and echocardiographic analyses (**f**) were performed. Asterisk, $P < 0.05$, two asterisks, $P < 0.005$ versus vehicle sham; dagger, $P < 0.05$ versus quinacrine sham; hash sign, $P < 0.05$ versus vehicle TAC. Error bars indicate s.e.m.; $n = 4$. **g**, Proposed mechanism underlying the transition from cardiac hypertrophy to heart failure.

(TUNEL) was significantly increased in the late phase of pressure overload, and this increase was attenuated by p53 deficiency (Supplementary Fig. 11a). Treatment with quinacrine increased the expression of p53 in both cardiomyocytes and endothelial cells but did not decrease the viability of these cells, microvessel density or cardiac function, in sham-operated mice (Fig. 4d, f, and Supplementary Figs 9 and 11b, c). In contrast, the number of TUNEL-positive endothelial cells and cardiomyocytes was significantly increased in the TAC heart treated with quinacrine (Supplementary Fig. 11b), which was associated with an increase in Bax expression (Supplementary Fig. 11c). Although p53-inducible pro-apoptotic genes have been implicated in hypoxia-mediated cell death[16,17], these results indicate that upregulated p53 may not cause cardiomyocyte death or systolic dysfunction directly but does so indirectly through the inhibition of cardiac angiogenesis.

Our results show that p53 is a crucial regulator in the induction of maladaptive hypertrophy, and that it does so by inhibiting cardiac angiogenesis (Fig. 4g). Gene therapy with constitutively active HIF-1α, which is resistant to degradation under normoxia, is currently being examined in clinical trials for peripheral vascular disease[18,19] and may be effective for the treatment of heart failure induced by pressure overload. Although further studies are required, inhibition of p53 or promotion of vascular growth in the heart may be a novel therapeutic strategy for preventing the transition from cardiac hypertrophy to heart failure. This strategy may be also useful to improve systolic dysfunction caused by various stimuli that increase cardiac p53 activity[20].

METHODS

Animal models. All protocols were approved by Chiba University review board. TAC was performed as described previously[21] on 8-week-old male C57BL/6 mice (SLC).

Physiological analysis and histological analysis. Echocardiography was performed as described previously[22]. Frozen cross-sections of heart samples were stained with antibodies against specific proteins as described in Supplementary Information.

Conditional ablation of the *Hif-1α* gene in cardiomyocytes of adult murine heart. We prepared transgenic mice in which a transgene encoding Cre recombinase fused to the mutated oestrogen receptor domains (MerCreMer) was driven by the cardiomyocyte-specific α-myosin heavy chain (MHC) promoter[23]. We then crossed the MHC-MerCreMer mice with mice that carried floxed *Hif-1α* alleles (*Hif-1α*$^{floxed/floxed}$)[24] and produced MHC-MerCreMer;*Hif-1α*$^{floxed/floxed}$ mutant mice.

Western blot analysis. Whole-cell lysates (30–50 μg) or nuclear extracts (10–20 μg) were resolved by SDS–polyacrylamide gel electrophoresis. Proteins were transferred to a polyvinylidene difluoride membrane (Millipore) and incubated with the first antibody followed by an anti-immunoglobulin-G–horseradish peroxidase antibody (Jackson ImmunoResearch). Specific proteins were detected by enhanced chemiluminescence (Amersham).

Statistical analysis. Data are shown as means ± s.e.m. Multiple group comparison was performed by one-way analysis of variance followed by the Bonferroni procedure for comparison of means.

Received 1 November 2006; accepted 15 January 2007.
Published online 4 March 2007.

1. Frey, N. & Olson, E. N. Cardiac hypertrophy: the good, the bad, and the ugly. *Annu. Rev. Physiol.* **65,** 45–79 (2003).
2. Levy, D. *et al.* Prognostic implications of echocardiographically determined left ventricular mass in the Framingham Heart Study. *N. Engl. J. Med.* **322,** 1561–1566 (1990).
3. Marcus, M. L. *et al.* Abnormalities in the coronary circulation that occur as a consequence of cardiac hypertrophy. *Am. J. Med.* **75,** 62–66 (1983).
4. Tomanek, R. J. Response of the coronary vasculature to myocardial hypertrophy. *J. Am. Coll. Cardiol.* **15,** 528–533 (1990).
5. Giordano, F. J. *et al.* A cardiac myocyte vascular endothelial growth factor paracrine pathway is required to maintain cardiac function. *Proc. Natl Acad. Sci. USA* **98,** 5780–5785 (2001).
6. Shyu, K. G. *et al.* Carvedilol prevents cardiac hypertrophy and overexpression of hypoxia-inducible factor-1α and vascular endothelial growth factor in pressure-overloaded rat heart. *J. Biomed. Sci.* **12,** 409–420 (2005).
7. Yoon, Y. S. *et al.* Progressive attenuation of myocardial vascular endothelial growth factor expression is a seminal event in diabetic cardiomyopathy: restoration of microvascular homeostasis and recovery of cardiac function in diabetic cardiomyopathy after replenishment of local vascular endothelial growth factor. *Circulation* **111,** 2073–2085 (2005).
8. Ingber, D. *et al.* Synthetic analogues of fumagillin that inhibit angiogenesis and suppress tumour growth. *Nature* **348,** 555–557 (1990).
9. Goldman, C. K. *et al.* Paracrine expression of a native soluble vascular endothelial growth factor receptor inhibits tumor growth, metastasis, and mortality rate. *Proc. Natl Acad. Sci. USA* **95,** 8795–8800 (1998).
10. Morani, A. *et al.* Lung dysfunction causes systemic hypoxia in estrogen receptor β knockout (ERβ$^{-/-}$) mice. *Proc. Natl Acad. Sci. USA* **103,** 7165–7169 (2006).
11. Semenza, G. L. Targeting HIF-1 for cancer therapy. *Nature Rev. Cancer* **3,** 721–732 (2003).
12. Pugh, C. W. & Ratcliffe, P. J. Regulation of angiogenesis by hypoxia: role of the HIF system. *Nature Med.* **9,** 677–684 (2003).
13. Blagosklonny, M. V. *et al.* p53 inhibits hypoxia-inducible factor-stimulated transcription. *J. Biol. Chem.* **273,** 11995–11998 (1998).
14. Ravi, R. *et al.* Regulation of tumor angiogenesis by p53-induced degradation of hypoxia-inducible factor 1α. *Genes Dev.* **14,** 34–44 (2000).
15. Gurova, K. V. *et al.* Small molecules that reactivate p53 in renal cell carcinoma reveal a NF-κB-dependent mechanism of p53 suppression in tumors. *Proc. Natl Acad. Sci. USA* **102,** 17448–17453 (2005).
16. Fei, P. *et al.* Bnip3L is induced by p53 under hypoxia, and its knockdown promotes tumor growth. *Cancer Cell* **6,** 597–609 (2004).
17. Kubasiak, L. A., Hernandez, O. M., Bishopric, N. H. & Webster, K. A. Hypoxia and acidosis activate cardiac myocyte death through the Bcl-2 family protein BNIP3. *Proc. Natl Acad. Sci. USA* **99,** 12825–12830 (2002).
18. Kelly, B. D. *et al.* Cell type-specific regulation of angiogenic growth factor gene expression and induction of angiogenesis in nonischemic tissue by a constitutively active form of hypoxia-inducible factor 1. *Circ. Res.* **93,** 1074–1081 (2003).
19. Patel, T. H. *et al.* Constitutively active HIF-1α improves perfusion and arterial remodeling in an endovascular model of limb ischemia. *Cardiovasc. Res.* **68,** 144–154 (2005).
20. Shizukuda, Y. *et al.* Targeted disruption of p53 attenuates doxorubicin-induced cardiac toxicity in mice. *Mol. Cell. Biochem.* **273,** 25–32 (2005).
21. Takimoto, E. *et al.* Sodium calcium exchanger plays a key role in alteration of cardiac function in response to pressure overload. *FASEB J.* **16,** 373–378 (2002).
22. Harada, M. *et al.* G-CSF prevents cardiac remodeling after myocardial infarction by activating the Jak-Stat pathway in cardiomyocytes. *Nature Med.* **11,** 305–311 (2005).
23. Sohal, D. S. *et al.* Temporally regulated and tissue-specific gene manipulations in the adult and embryonic heart using a tamoxifen-inducible Cre protein. *Circ. Res.* **89,** 20–25 (2001).
24. Tomita, S. *et al.* Defective brain development in mice lacking the Hif-1α gene in neural cells. *Mol. Cell. Biol.* **23,** 6739–6749 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Fujita, R. Kobayashi and M. Ikeda for technical support. This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas and for Exploratory Research, Ministry of Education, Culture, Sports, Science and Technology; Health and Labour Sciences Research Grants; Research on Measures for Intractable Diseases; Grants from Goho Life Sciences International Fund; an Academic Award of the Mochida Memorial Foundation and Uehara Memorial Foundation (to I.K.); and grants from the Suzuken Memorial Foundation, the NOVARTIS Foundation and the Ministry of Education, Culture, Sports, Science and Technology of Japan (to T.M.).

Author Contributions M.S., T.M., H.T., H.M., M.O., Y.Q., H.A., K.T., Y.K., M.H., I.S. and Y.Z. performed the experiments. T.A., H.H., S.T. and J.D.M. provided reagents or mice. M.S., T.M. and I.K. designed and prepared the manuscript. I.K. planned and supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.K. (komuro-tky@umin.ac.jp).

LETTERS

BluB cannibalizes flavin to form the lower ligand of vitamin B_{12}

Michiko E. Taga[1]*, Nicholas A. Larsen[2]*, Annaleise R. Howard-Jones[2], Christopher T. Walsh[2] & Graham C. Walker[1]

Vitamin B_{12} (cobalamin) is among the largest known non-polymeric natural products, and the only vitamin synthesized exclusively by microorganisms[1]. The biosynthesis of the lower ligand of vitamin B_{12}, 5,6-dimethylbenzimidazole (DMB), is poorly understood[1–3]. Recently, we discovered that a *Sinorhizobium meliloti* gene, *bluB*, is necessary for DMB biosynthesis[4]. Here we show that BluB triggers the unprecedented fragmentation and contraction of the bound flavin mononucleotide cofactor and cleavage of the ribityl tail to form DMB and D-erythrose 4-phosphate. Our structural analysis shows that BluB resembles an NAD(P)H-flavin oxidoreductase, except that its unusually tight binding pocket accommodates flavin mononucleotide but not NAD(P)H. We characterize crystallographically an early intermediate along the reaction coordinate, revealing molecular oxygen poised over reduced flavin. Thus, BluB isolates and directs reduced flavin to activate molecular oxygen for its own cannibalization. This investigation of the biosynthesis of DMB provides clarification of an aspect of vitamin B_{12} that was otherwise incomplete, and may contribute to a better understanding of vitamin B_{12}-related disease.

The lower ligand of vitamin B_{12}, DMB, is the only component of B_{12} for which no biosynthetic enzymes have been identified[1–3]. Previous studies using cell extracts have demonstrated the oxygen-dependent transformation of flavin mononucleotide (FMN) to DMB (Fig. 1a); however, the mechanism of conversion and enzymes involved in this pathway remain unknown[3,5,6]. Recently, a DMB auxotroph of the symbiotic nitrogen-fixing bacterium *Sinorhizobium meliloti*was identified[4]. This mutant, *bluB*, exhibited abnormal exopolysaccharide structure, as indicated by a calcofluor-bright phenotype, and was deficient in free-living growth and symbiosis with its plant host. A putative B_{12}-binding riboswitch immediately upstream of the *bluB* gene indicated a B_{12} biosynthetic role[7]; addition of vitamin B_{12} or DMB reversed all of the *bluB* mutant phenotypes[4]. BluB homologues exist in phylogenetically and metabolically diverse prokaryotic genomes, including notable pathogens of the genera

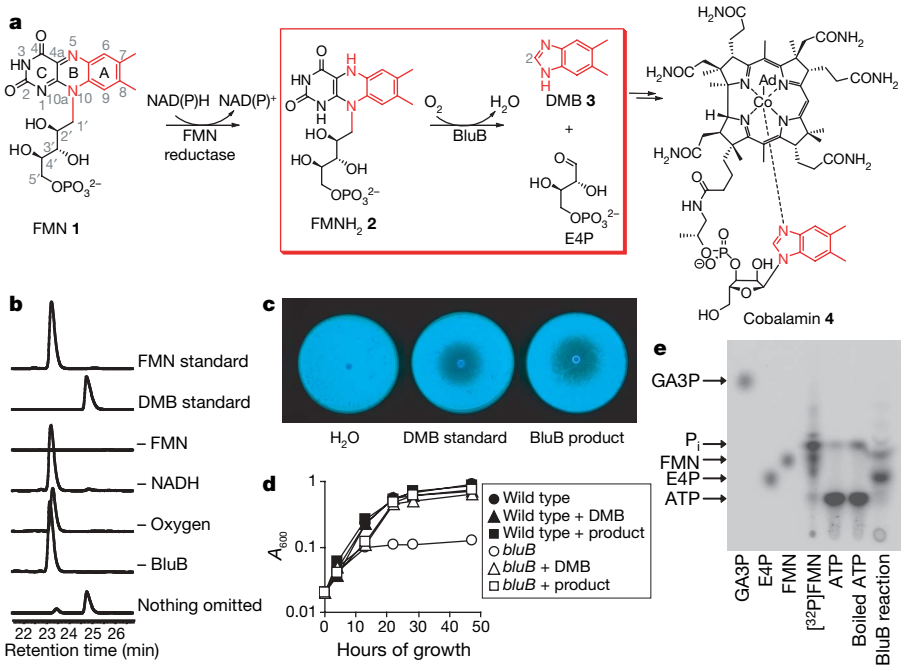


Figure 1 | BluB catalyses DMB production. a, Overall scheme for DMB biosynthesis and B_{12} formation. The atoms converted to DMB are shown in red; the reaction catalysed by BluB is in the red box. **b**, HPLC traces (280 nm) of reactions containing 10 μM BluB, 20 mM NADH and 100 μM FMN, or lacking the indicated components. Similar conversion was observed in reactions containing NADPH instead of NADH (not shown). In the absence of NAD(P)H, trace amounts of product formed, probably through adventitious photoreduction of FMN followed by turnover in the active site. **c**, Calcofluor fluorescence of the *bluB* mutant. **d**, Growth of wild-type *S. meliloti* and the *bluB* mutant. **e**, Radio-TLC of the BluB reaction containing [^{32}P]FMN. Arrows represent migration of standards. The same result was obtained using two additional eluant systems (data not shown).

[1]Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. [2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work.

Table 1 | Kinetic parameters of BluB-catalysed DMB synthesis

Compound tested	k_{cat} (apparent) (h^{-1})*	K_m (apparent) (μM)*	K_i (μM)*
FMN†	15 ± 3	64 ± 23	231 ± 83
NADH‡	7.2 ± 0.7	5,100 ± 1,700	NA
NADPH‡	7.6 ± 0.5	4,400 ± 1,100	NA

k_{cat}, catalytic rate constant; K_i, inhibition constant; NA, not applicable.
* Initial velocity data for DMB formation were fitted to the Michaelis–Menten equation using KaleidaGraph (Synergy Software). Parameters for FMN were calculated using a modified equation that takes into account substrate inhibition (Supplementary Fig. 2). Errors represent standard deviation.
† Reactions contained 10 μM BluB, 40 mM NADH and FMN at concentrations ranging from 10 μM to 2 mM.
‡ Reactions contained 10 μM BluB, 200 μM FMN and NAD(P)H at concentrations ranging from 1 to 40 mM.

Mycobacterium, *Brucella* and *Vibrio*, as well as the eukaryotic mouse malarial agent *Plasmodium yoelii*, indicating a ubiquitous role in DMB synthesis[4,7].

The protein sequence of BluB has ~16% identity to several enzymes of the nitroreductase/flavin oxidoreductase family. These enzymes catalyse the NAD(P)H-dependent reduction of flavin for diverse redox reactions[8,9]. On the basis of this sequence similarity and our previous findings, we proposed that BluB might catalyse the transformation of FMN to DMB. Indeed, high-performance liquid chromatography (HPLC) analysis reveals the BluB-dependent consumption of FMN and formation of DMB, as confirmed by co-elution, ultraviolet/visible spectroscopy, ^{1}H NMR spectroscopy and mass spectrometry, by comparison with an authentic standard (Fig. 1b; see also Supplementary Fig. 1). Furthermore, two of the phenotypes originally observed for the *bluB* mutant—the calcofluor fluorescence and poor growth in minimal media—are also reversed by the BluB reaction product (Fig. 1c, d). The catalysed reaction requires oxygen, consistent with previous *ex vivo* data[3,5,10] (Fig. 1b). BluB also consumes $FMNH_2$ as a substrate in the absence of NAD(P)H (data not shown). Thus, $FMNH_2$ is the true substrate for BluB, whereas molecular oxygen drives oxidative fragmentation of the heterocycle. NAD(P)H is only required initially to reduce FMN.

Kinetic characterization of BluB reveals a non-Michaelis–Menten profile, as the reaction is inhibited by high concentrations of FMN (Supplementary Fig. 2)[11,12]. The apparent K_m for FMN (64 μM) is close to its physiological concentration, whereas inhibition occurs at non-physiological concentrations[13] (Table 1). The apparent K_m values for NADH and NADPH (5.1 and 4.4 mM, respectively) are 10–100-fold higher than their physiological concentrations and imply an unfavourable reductase activity[13]. These data suggest an alternative kinetic model in which $FMNH_2$ is the substrate and NAD(P)H does not saturate the active site, but rather reduces FMN in solution.

BluB, unlike the closely related oxidoreductase enzyme family, does not catalyse NAD(P)H-dependent FMN reduction, whereas *Escherichia coli* SsuE, a well-characterized flavin reductase[11], accelerates FMN reduction by at least 60-fold above background levels (Supplementary Table 1). Significant non-enzymatic reduction of FMN by NAD(P)H (Supplementary Table 1) enables BluB to turn over *in vitro*, whereas an accessory reductase probably delivers $FMNH_2$ to BluB *in vivo*. Indeed, addition of SsuE to the BluB reaction increases the initial rate ~20-fold (Supplementary Table 2). The rate of DMB formation increases markedly ($k_{obs} = 3\ min^{-1}$) in reactions containing SsuE and a concentration of FMN (1 μM) below BluB's dissociation constant ($K_d \approx 2$ μM) (Supplementary Fig. 3). These data suggest that, despite its apparent sequence homology, BluB is not a flavin reductase but may require a separate reductase for delivery of $FMNH_2$ (ref. 9).

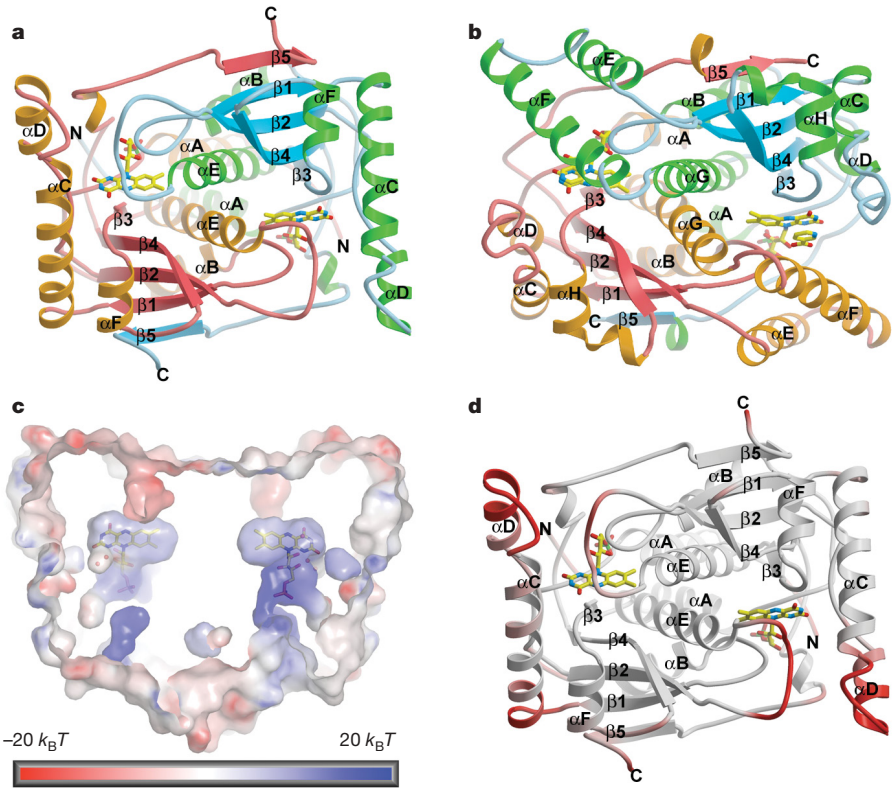


Figure 2 | Structure of BluB. **a**, Ribbon diagram of BluB with FMN in the binding pocket (stick representation). The two-fold axis is perpendicular to the plane of the figure. **b**, *E. coli* nitroreductase NfsB (Protein Data Bank 1ICR)[21] with FMN and nicotinic acid (stick representation) in the binding pocket. **c**, Cross-section of BluB's molecular surface. The two-fold axis lies along the *y* axis such that the *si*-face of FMN is viewed on the left and *re*-face on the right. The surface is coloured according to electrostatic potential, where blue is electropositive, red is electronegative and k_B is Boltzmann's constant. The back and front of the surface are cut away to reveal the FMN binding pocket buried in the dimer interface. The pocket wraps snugly around FMN, preventing interaction with other substrates. **d**, The BluB ribbon diagram has been coloured according to B-factor. Red represents flexible regions that may control or gate access to the active site.

The remarkable contraction of the isoalloxazine ring of FMN to form DMB requires cleavage of the ribityl tail. Numerous studies have demonstrated that carbon C2 of DMB is derived from carbon C1′ of FMN[3,6,14,15], and previously proposed mechanisms have suggested that the remainder of the ribityl moiety of FMN is released as either D-glyceraldehyde 3-phosphate (GA3P) and CO_2 (refs 3,16) or D-erythrose 4-phosphate (E4P)[10,17]. Thin-layer chromatography (TLC) analysis of reactions containing [^{32}P]FMN shows that the product mixture contains species co-migrating with E4P and FMN, but not GA3P, indicating that E4P is the major phosphorous-containing product (Fig. 1e). ^{31}P-NMR confirms that E4P is formed, whereas GA3P is not detectable (Supplementary Fig. 4). Therefore, BluB catalyses the elimination of the ribityl tail of FMN by cleavage between C1′ and C2′ to form the four-carbon sugar E4P.

The crystal structure of BluB resembles the NAD(P)H-flavin oxidoreductases, which include flavin reductase (FRP) of *Vibrio harveyi* and *Vibrio fischeri*[18,19], NADH oxidase (NOX) of *Thermus thermophilus*[20], and the nitroreductases NfsA and NfsB of *E. coli*[21] (Fig. 2a, b; see also Supplementary Fig. 5). The root mean square (r.m.s.) deviation between these enzymes and BluB is 2.1–3.8 Å over 60–80% of the structure, depending on the family member (Fig. 2a, b)[22]. Comparison of BluB to these structural homologues reveals a unique insertion in αC that significantly extends the αC-loop-αD motif to form a lid that completely buries FMN (Fig. 2a–c). In addition, there is a unique deletion after strand β2, replacing two helices (αE and αF in the typical reductase) with a short loop that further constricts the binding pocket (Fig. 2a, b; see also Supplementary Fig. 5). These unique elements appear to have flexibility, and thus provide gated access to the active site (Fig. 2d). Nevertheless, they conspire to isolate FMN from other substrates in an extremely snug, electropositive pocket (Fig. 2c; see also Supplementary Fig. 6). In contrast, the open, solvent-accessible active site in oxidoreductases facilitates promiscuous exchange of substrates (Fig. 2b)[19,21]. Thus, the active site of BluB carefully excludes interaction with any substrate other than dioxygen.

We obtained the crystal structure of BluB bound to substrate ($FMNH_2$) and molecular oxygen by soaking crystals in the reducing agent dithionite and then back-soaking in oxygenated mother liquor. The slow turnover of BluB (Table 1) is retarded further because of the acidic crystallization buffer, enabling capture of a crystallographic snap-shot along the reaction coordinate. Notably, the dithionite-soaked crystals turn purple, probably representing a non-physiological, charge-transfer complex with the reduced flavin anion (Supplementary Fig. 7). On reintroduction of oxygen, the purple crystals become pale yellow or clear, with the spectroscopic signature of $FMNH_2$ (Supplementary Fig. 7). These clear crystals disintegrate rapidly; evidently, turnover in the crystal is coupled to a structural rearrangement that disrupts the lattice. We successfully froze one such exceptionally large, pale-yellow crystal and collected diffraction data to 2.9 Å. The conformational difference between oxidized and reduced FMN is subtle (Fig. 3a, b). Notably, some of the H-bonds rearrange at the N1 position, which is protonated in the reduced form at acidic pH (Fig. 3; see also Supplementary Table 3). Asp 32 also shows conformational mobility, and is able to form a close contact with C1′ of the ribityl tail and N1 (Fig. 3c, d; see also Supplementary Table 3). This residue is conserved in BluB homologues[4], but is not observed in oxidoreductases. Mutation of Asp 32 to Ala, Asn or Ser abrogates DMB formation but retains flavin binding, indicating that it may be essential for catalysis (Fig. 4).

A strong electron density peak appears over the *re*-face of reduced flavin, replacing a weaker peak (modelled as water) in the oxidized structure (Fig. 3c, d). Modelling this strong peak as water results in unacceptably low B-factors, whereas refinement as molecular oxygen accounts for the extra density and is consistent with the enzyme's dependence on oxygen (Fig. 3d; see also Supplementary Table 4).

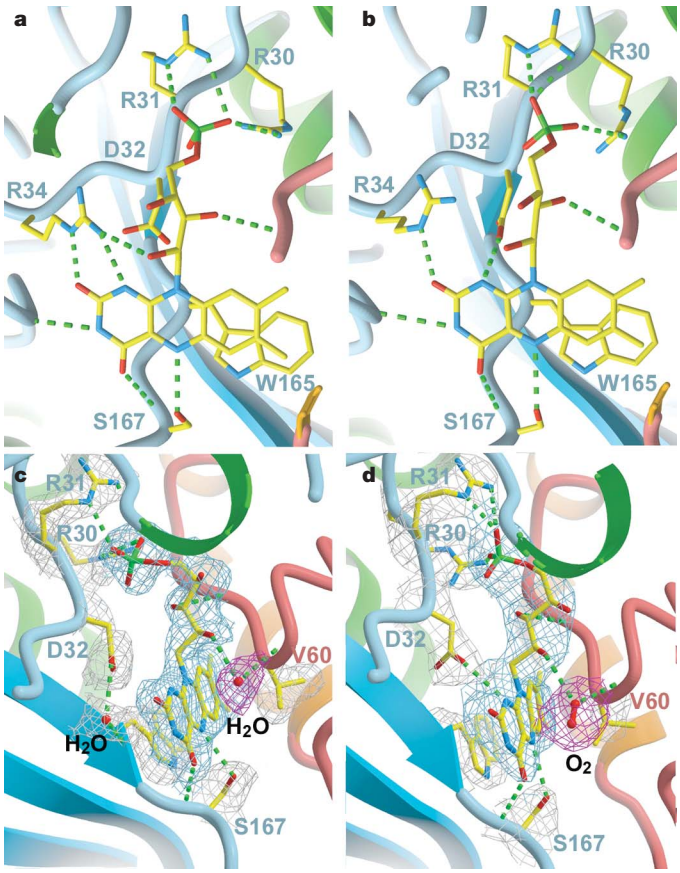


Figure 3 | Active site of BluB. **a**, **b**, The active site with oxidized FMN (**a**) and reduced FMN (**b**), viewed from the *re*-face. H-bonds are represented as dashed lines. For clarity, water molecules are not rendered in this view. The rearrangement in H-bonds around N1 reflects the change in protonation in the reduced structure. Asp 32 may also form a close contact with C1′ of the ribityl chain, suggesting a potential catalytic role for this residue. **c**, **d**, Side views of the active site in the oxidized (**c**) and reduced (**d**) structures. For clarity, Arg 34 has not been rendered in this view. The sigma-A weighted $2F_o-F_c$ electron density map is contoured at 1σ and coloured grey (around protein side chains), blue (around FMN/$FMNH_2$) and red (around water/oxygen) to enhance contrast.

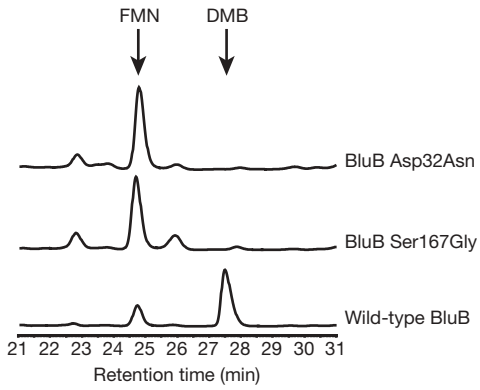


Figure 4 | Mutation of Asp 32 or Ser 167 eliminates DMB production. HPLC traces (280 nm) of reactions performed with BluB proteins containing the point mutations Asp32Asn or Ser167Gly. The peaks corresponding to FMN and DMB, as confirmed by retention time and ultraviolet/visible spectra, are labelled. No DMB is detected in the Asp32Asn mutant; DMB produced by the Ser167Gly mutant is ~3% of the wild-type level. No DMB is detected in reactions containing BluB with Asp32Ala, Asp32Ser, or Ser167Cys point mutations (not shown). The size of the small peaks at 22.7 and 25.8 min in the reactions with mutant enzymes is reduced by addition of catalase, indicating that H_2O_2 is generated in reactions with mutant BluB, probably via futile cycling through a peroxyflavin intermediate (not shown).

This oxygen is poised for attack at C4a and anchored by two H-bonds from the O2′ hydroxyl and the backbone amide from residue 61. These H-bonds might serve as a 'peroxyanion hole' for stabilization of a reactive peroxyflavin intermediate, analogous to the oxyanion hole that stabilizes the tetrahedral intermediate in serine proteases and hydrolases[23] (Fig. 3d). Such stabilization would deter uncoupling of hydrogen peroxide (H_2O_2), which occurs favourably in solution. Mono-oxygenases direct formation of a peroxyflavin intermediate for subsequent hydroxylation reactions, and also seem to use peroxyanion holes, which masquerade as two backbone amides, Arg or Asn[24–26]. Finally, the peroxyanion hole of BluB is defined, in part, by the orientation of the ribityl tail, which resembles that in oxidoreductases, but differs from mono-oxygenases, where the tail is swivelled away and sheltered from the site of oxygen attack[24–27]. Thus, BluB is a hybrid enzyme with structural similarity to the flavin oxidoreductases and functional relatedness to the mono-oxygenases.

BluB has additional unique features observed in neither the oxidoreductases nor the mono-oxygenases that may have critical functions in directing DMB formation. Asp 32 and Ser 167 are completely conserved in BluB homologues but have not been observed in oxidoreductases or mono-oxygenases[8,19,21,24–27]. Ser 167 forms an H-bond to N5 of flavin, whereas H-bonds to N5 are fulfilled by backbone amides in oxidoreductases and are seldom observed in mono-oxygenases. Consistent with a critical catalytic role, mutation of Ser 167 to Gly reduces DMB production ~30-fold, and mutation to Cys completely abolishes DMB formation (Fig. 4). Moreover, the H-bond to N5 in oxidoreductases typically forms from one face of the flavin, with the opposite face left unhindered for interaction with substrate and oxygen[8,19,21]. In contrast, BluB appears to donate this H-bond and bind oxygen from the same face (Fig. 3c, d). This stereochemistry may serve as an additional control mechanism to prevent unwanted reactions from all but the smallest possible substrates, namely molecular oxygen.

DMB formation involves controlled breakage of three bonds and formation of one, facilitating contraction of the central six-membered ring of $FMNH_2$ to form the five-membered ring of DMB. A mechanistic proposal for this transformation must incorporate the following findings of this study: (1) requirement for molecular oxygen; (2) rupture of the ribityl side chain between C1′ and C2′ to yield E4P; (3) absence of a requirement for NAD(P)H; and (4) the critical roles of Asp 32 and Ser 167 in catalysis. Furthermore, earlier observations using cell extracts showed that C2 of DMB is derived from C1′ of FMN[6,14,16], and that urea is produced from this reaction[3]. At this point, it is not possible to provide a detailed mechanism for this transformation. Rather, we postulate a framework to explain the fragmentation and reconstruction of $FMNH_2$ to form DMB (see 'Discussion' and 'Scheme 1' in Supplementary Information). The closed active site strongly indicates an intramolecular route for the overall reaction, requiring not only a difficult chemical transformation but also major conformational change. Such rearrangement of the protein structure is a probable explanation for the observed fragmentation of BluB crystals on introduction of dioxygen.

BluB is thus the first characterized enzyme responsible for the formation of DMB, hence elucidating the last unknown step in the biosynthesis of B_{12}. Moreover, this investigation has revealed a novel enzymatic activity in the fragmentation of a flavin coenzyme. Destruction of one cofactor to form another is unusual, but has parallels with the cannibalization of iron–sulphur clusters in the biosynthesis of biotin (vitamin H) and lipoamide coenzymes[28,29]. The biochemical and structural novelty of BluB and the widespread distribution of BluB homologues throughout both prokaryotic kingdoms call for a new enzyme classification. Therefore, we propose that this enzyme belongs to a new 'flavin destructase' family.

METHODS

The procedures for protein expression, purification, crystallization, structure determination and refinement are described in Supplementary Methods.

Enzymatic reactions. For endpoint assays, reactions were assembled in an anaerobic glove box, exposed to ambient O_2 (except where indicated), and quenched with 6.4% trichloroacetic acid after overnight incubation at room temperature. Reaction buffer contained 50 mM HEPES pH 7.5 and 1 mM dithiothreitol. Two-hundred-microlitre samples were analysed by RP-HPLC with a C18 column (Beckman Coulter) (11% acetonitrile, 0.1% TFA in water, 1 ml min^{-1}), monitoring absorbance at 280 and 450 nm. The peak identities were confirmed using a coupled ultraviolet/visible diode array detector. Standards contained 100 μM FMN or DMB. For rate measurements, 250 μl aliquots from 500 μl reactions were quenched at two different time points, and reaction rates for the two time points were averaged. DMB concentrations were calculated based on integrated peak areas at 280 nm compared to a standard curve. His-tagged SsuE protein was purified as described[11].

Biological assays. For calcofluor analysis, 10 μl of 10 μM DMB standard, BluB reaction product, or water was added to a sterile filter disk on LB-calcofluor plates containing a lawn of an *S. meliloti bluB* mutant strain[4]. For growth measurements, M9 medium containing cobalt[4] with nothing added, 100 nM DMB standard, or 100 nM BluB reaction product was inoculated with wild-type *S. meliloti* strain Rm1021 or the *S. meliloti bluB* mutant, and absorbance at 600 nm (A_{600}) was monitored over time.

Radio-TLC. The procedure for [^{32}P]FMN production by riboflavin kinase is described in Supplementary Methods. Reactions containing 10 μM [^{32}P]FMN, 5 mM NADPH, 1 μM SsuE and 10 μM BluB were incubated for 2 h at 30 °C and quenched by passing the reactions through 30,000 MWCO filters (Microcon). The reactions were spotted on cellulose-TLC plates and developed with 3:1:6 *n*-butanol:acetic acid:water. The unlabelled GA3P and E4P standards included on the plate were stained with ceric ammonium molybdate, and ~1 nCi [γ-^{32}P]ATP was overlaid onto the stained spots. Plates were visualized by phosphorimaging.

Received 25 October 2006; accepted 19 January 2007.

1. Roth, J. R., Lawrence, J. G. & Bobik, T. A. Cobalamin (coenzyme B_{12}): synthesis and biological significance. *Annu. Rev. Microbiol.* **50,** 137–181 (1996).
2. Warren, M. J. Finding the final pieces of the vitamin B_{12} biosynthetic jigsaw. *Proc. Natl Acad. Sci. USA* **103,** 4799–4800 (2006); published online 27 March 2006.
3. Renz, P. in *Chemistry and Biochemistry of B_{12}* (ed. Banerjee, R.) 557–575 (John Wiley & Sons, New York, 1999).
4. Campbell, G. R. O. *et al. Sinorhizobium meliloti bluB* is necessary for production of 5,6-dimethylbenzimidazole, the lower ligand of B_{12}. *Proc. Natl Acad. Sci. USA* **103,** 4634–4639 (2006).
5. Renz, P. Riboflavin as precursor in the biosynthesis of the 5,6-dimethylbenzimidazole-moiety of vitamin B_{12}. *FEBS Lett.* **6,** 187–189 (1970).
6. Renz, P. & Weyhenmeyer, R. Biosynthesis of 5,6-dimethylbenzimidazole from riboflavin. Transformation of C-1′ of riboflavin into C-2 of 5,6-dimethylbenzimidazole. *FEBS Lett.* **22,** 124–126 (1972).
7. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A. & Gelfand, M. S. Comparative genomics of the vitamin B_{12} metabolism and regulation in prokaryotes. *J. Biol. Chem.* **278,** 41148–41159 (2003); published online 17 July 2003.
8. Fraaije, M. W. & Mattevi, A. Flavoenzymes: diverse catalysts with recurrent features. *Trends Biochem. Sci.* **25,** 126–132 (2000).
9. Tu, S. C. Reduced flavin: donor and acceptor enzymes and mechanisms of channeling. *Antioxid. Redox Signal.* **3,** 881–897 (2001).
10. Horig, J. A. & Renz, P. Biosynthesis of vitamin B_{12}. Some properties of the 5,6-dimethylbenzimidazole-forming system of *Propionibacterium freudenreichii* and *Propionibacterium shermanii*. *Eur. J. Biochem.* **105,** 587–592 (1980).
11. Eichhorn, E., van der Ploeg, J. R. & Leisinger, T. Characterization of a two-component alkanesulfonate monooxygenase from *Escherichia coli*. *J. Biol. Chem.* **274,** 26639–26646 (1999).
12. Lei, B. & Tu, S. C. Mechanism of reduced flavin transfer from *Vibrio harveyi* NADPH-FMN oxidoreductase to luciferase. *Biochemistry* **37,** 14623–14629 (1998).
13. Bochner, B. R. & Ames, B. N. Complete analysis of cellular nucleotides by two-dimensional thin layer chromatography. *J. Biol. Chem.* **257,** 9759–9769 (1982).
14. Lingens, B., Schild, T. A., Vogler, B. & Renz, P. Biosynthesis of vitamin B_{12}. Transformation of riboflavin 2H-labeled in the 1′R position of 1′S position into 5,6-dimethylbenzimidazole. *Eur. J. Biochem.* **207,** 981–985 (1992).
15. Keck, B., Munder, M. & Renz, P. Biosynthesis of cobalamin in *Salmonella typhimurium*: transformation of riboflavin into the 5,6-dimethylbenzimidazole moiety. *Arch. Microbiol.* **171,** 66–68 (1998).
16. Kolonko, B., Horig, J. A. & Renz, P. Transformation of [5′-^{3}H]riboflavin into 5,6-dimethylbenzimidazole. *Z. Naturforsch. C* **47,** 171–176 (1992).
17. Maggio-Hall, L. A., Dorrestein, P. C., Escalante-Semerena, J. C. & Begley, T. P. Formation of the dimethylbenzimidazole ligand of coenzyme B_{12} under physiological conditions by a facile oxidative cascade. *Org. Lett.* **5,** 2211–2213 (2003).
18. Tanner, J. T., Lei, B., Tu, S. C. & Krause, K. L. Flavin reductase P: structure of a dimeric enzyme that reduces flavin. *Biochemistry* **35,** 13531–13539 (1996).
19. Koike, H. *et al.* 1.8 Å crystal structure of the major NAD(P)H:FMN oxidoreductase of a bioluminescent bacterium, *Vibrio fischeri*: overall structure, cofactor and

substrate-analog binding, and comparison with related flavoproteins. *J. Mol. Biol.* **280,** 259–273 (1998).
20. Hecht, H. J., Erdmann, H., Park, H. J., Sprinzl, M. & Schmid, R. D. Crystal structure of NADH oxidase from *Thermus thermophilus*. *Nature Struct. Mol. Biol.* **2,** 1109–1114 (1995).
21. Lovering, A. L., Hyde, E. I., Searle, P. F. & White, S. A. The structure of *Escherichia coli* nitroreductase complexed with nicotinic acid: three crystal forms at 1.7 Å, 1.8 Å and 2.4 Å resolution. *J. Mol. Biol.* **309,** 203–213 (2001).
22. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233,** 123–138 (1993).
23. Larsen, N. A., Lin, H., Wei, R., Fischbach, M. A. & Walsh, C. T. Structural characterization of enterobactin hydrolase IroE. *Biochemistry* **45,** 10184–10190 (2006).
24. Eswaramoorthy, S., Bonanno, J. B., Burley, S. K. & Swaminathan, S. Mechanism of action of a flavin-containing monooxygenase. *Proc. Natl Acad. Sci. USA* **103,** 9832–9837 (2006).
25. Gatti, D. L., Entsch, B., Ballou, D. P. & Ludwig, M. L. pH-dependent structural changes in the active site of *p*-hydroxybenzoate hydroxylase point to the importance of proton and water movements during catalysis. *Biochemistry* **35,** 567–578 (1996).
26. Malito, E., Alfieri, A., Fraaije, M. W. & Mattevi, A. Crystal structure of a Baeyer–Villiger monooxygenase. *Proc. Natl Acad. Sci. USA* **101**, 13157–13162 (2004); published online 24 August 2004.
27. Dong, C. *et al.* Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science* **309,** 2216–2219 (2005).
28. Berkovitch, F., Nicolet, Y., Wan, J. T., Jarrett, J. T. & Drennan, C. L. Crystal structure of biotin synthase, an *S*-adenosylmethionine-dependent radical enzyme. *Science* **303,** 76–79 (2004).
29. Cicchillo, R. M. & Booker, S. J. Mechanistic investigations of lipoic acid biosynthesis in *Escherichia coli*: both sulfur atoms in lipoic acid are contributed by the same lipoyl synthase polypeptide. *J. Am. Chem. Soc.* **127,** 2860–2861 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by NIH grants to G.C.W. and C.T.W. and postdoctoral fellowships from the Jane Coffin Childs Memorial Fund for Medical Research to M.E.T. and N.A.L. G.C.W. is an American Cancer Society Research Professor. We are grateful to E. Yeh for providing purified SsuE and H. Zhang for purified riboflavin kinase. We thank C. Sheahan and G. Heffron for their assistance with ^{31}P-NMR, and A. Haykov for assistance with protein purification. We acknowledge the Advanced Light Source for beam time. We thank S. Harrison, T. Begley, C. Drennan and members of the Walsh and Walker laboratories for helpful discussions.

Author Information The coordinates and structure factors for BluB-FMN, BluB-FMNA (flavin anion) and BluB-FMNH$_2$ have been deposited in the Protein Data Bank under accession codes 2ISJ, 2ISK and 2ISL, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.C.W. (gwalker@mit.edu) or C.T.W. (christopher_walsh@hms.harvard.edu).

Vol 446|22 March 2007|doi:10.1038/nature05625

LETTERS

Peptide bond formation destabilizes Shine–Dalgarno interaction on the ribosome

Sotaro Uemura[1], Magdalena Dorywalska[2], Tae-Hee Lee[1], Harold D. Kim[1], Joseph D. Puglisi[2] & Steven Chu[1,3,4]

The ribosome is a molecular machine that translates the genetic code contained in the messenger RNA into an amino acid sequence through repetitive cycles of transfer RNA selection, peptide bond formation and translocation[1–3]. Here we demonstrate an optical tweezer assay to measure the rupture force between a single ribosome complex and mRNA. The rupture force was compared between ribosome complexes assembled on an mRNA with and without a strong Shine–Dalgarno (SD) sequence—a sequence found just upstream of the coding region of bacterial mRNAs, involved in translation initiation[4,5]. The removal of the SD sequence significantly reduced the rupture force in complexes carrying an aminoacyl tRNA, Phe-tRNA^{Phe}, in the A site, indicating that the SD interactions contribute significantly to the stability of the ribosomal complex on the mRNA before peptide bond formation. In contrast, the presence of a peptidyl tRNA analogue, N-acetyl-Phe-tRNA^{Phe}, in the A site, which mimicked the post-peptidyl transfer state, weakened the rupture force as compared to the complex with Phe-tRNA^{Phe}, and the resultant force was the same for both the SD-containing and SD-deficient mRNAs. These results suggest that formation of the first peptide bond destabilizes the SD interaction, resulting in the weakening of the force with which the ribosome grips an mRNA. This might be an important requirement to facilitate movement of the ribosome along mRNA during the first translocation step.

The ribosome is a complex catalytic machine, composed of more than 50 different proteins and three RNA chains, that performs protein synthesis. The 30S small subunit provides a framework on which tRNAs can accurately pair with the mRNA codons, whereas the 50S large subunit catalyses the formation of peptide bonds[1–3]. Structural and biochemical data have revealed a network of contacts between tRNAs bound to adjacent codons in the peptidyl- and aminoacyl-tRNA sites (P and A sites) and both the 30S and 50S subunits[6–9]. Ribosome–tRNA–mRNA interactions are required for the maintenance and regulation of the ribosomal complex stability during all stages of translation, yet processive translocation requires relaxation of the interactions for movement of tRNA–codon complexes with respect to the ribosome. Initial positioning of the ribosome on mRNA involves the recognition of a purine-rich sequence, known as the Shine–Dalgarno (SD) sequence, located upstream of the AUG initiation codon on the mRNA and complementary to the 3′ end of the 16S rRNA[4,5].

Although the detailed interactions between the translation components have been revealed by extensive structural and biochemical studies, it is not known how the precise regulation of macromolecular movements required at each stage of translation is achieved. It remains unclear what signal induces the substantial and well-tuned macromolecular forces that the ribosome must generate following every peptide bond formation to trigger the progress of its 25-kDa tRNA substrates through the intersubunit active sites along with its own precise directional movement by one codon down the mRNA.

Here, using optical tweezers, we measure directly the forces exerted between the ribosome and mRNA in the context of various tRNAs before and after peptide bond formation. Ribosome complexes were assembled on a 57 nucleotide mRNA derived from the T4 gene 32

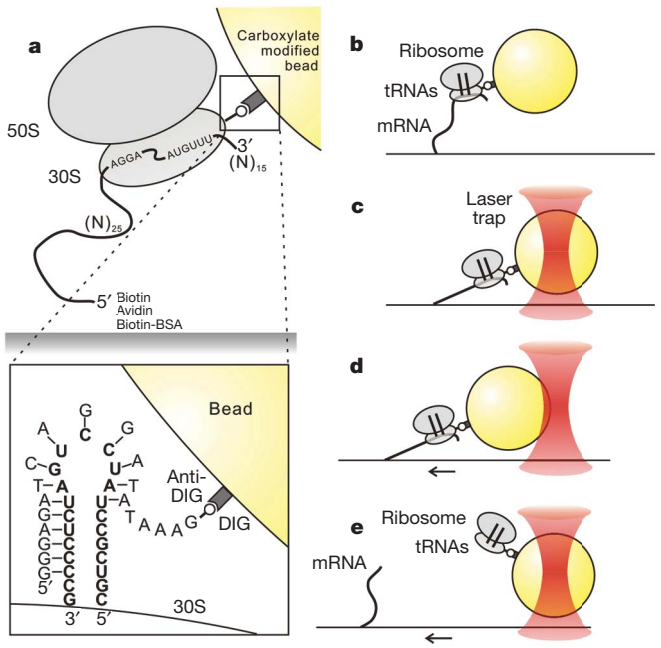


Figure 1 | Experimental design for rupture force measurements on the ribosome. a, The molecular attachments within the mRNA–ribosome–bead complex. Ribosomal particles were assembled on a short mRNA tethered to the surface via biotin–streptavidin linkage. A digoxigenin-modified oligonucleotide was designed to hybridize to an rRNA loop extension on the small ribosomal subunit (see magnified view at bottom). A bead coated with anti-digoxigenin antibody was conjugated to the oligonucleotide and used for optical trapping of the ribosomal complex. The sequence 'AGGA' on the mRNA designates the strong SD region, 'AUGUUU' indicates the first two codons of the mRNA, $(N)_{25}$ and $(N)_{15}$ specify the number of residues in the upstream and downstream regions, respectively, and 'Biotin Avidin Biotin-BSA' indicates the details of mRNA attachment to the surface. **b**, The tethered ribosome–bead complex fluctuates around the point of surface attachment. The two parallel lines within the ribosome signify the two tRNAs bound at the A and P sites. **c**, Optical tweezers are used to trap the bead. **d**, As the stage with the attached ribosome–bead complex is moved in one direction, the force exerted on the complex increases and the bead becomes displaced. **e**, Eventually the external force becomes sufficient to rupture the complex, and the bead returns to the trap centre position.

[1]Department of Physics, [2]Department of Structural Biology, Stanford University, Stanford, California 94305, USA. [3]Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [4]Departments of Physics, Molecular and Cellular Biology, University of California, Berkeley, California 94720, USA.

mRNA, containing a natural SD sequence and a 5′ biotin modification to tether the complex to streptavidin-derivatized quartz surfaces. For optical trapping, 1-μm-diameter carboxylate-modified beads coated with anti-digoxigenin antibody were bound to a digoxigenin-modified DNA oligonucleotide designed to hybridize to an extension genetically engineered into helix 44 of the *Escherichia coli* 16S rRNA[10] (Fig. 1a, b). A single bead bound to a single ribosome complex—as verified by single step fluorescence photobleaching of Cy3-labelled tRNA in the P site (Supplementary Fig. 1a, b)—was tethered to the surface via mRNA and trapped with optical tweezers (Fig. 1c). The piezo stage was moved at a constant velocity ($\sim$100 nm s^{-1}) in one direction followed by a movement of the bead (Fig. 1d) until the external force exerted on the ribosome resulted in a rupture event and a rapid return of the bead to the trap centre position (Fig. 1e). External force was applied at a relatively fast rate to minimize possible dissociation of the ribosome complex during the measurement (Supplementary Methods).

Several examples of rupture force events recorded for tethered mRNA–ribosome–bead complexes are shown in Fig. 2a. Rupture events probably result from ribosome dissociation from the mRNA, as no rupture events were observed in the following two sets of control experiments: in one set of experiments, the ribosomes were crosslinked directly to a cysteine-reactive surface instead of being tethered via mRNA (Fig. 2b); in the other, the ribosome–mRNA complex was replaced by a surface-attached biotinylated RNA oligonucleotide of the same sequence as the rRNA extension complementary to the DNA oligonucleotide–bead conjugate (Fig. 2c). These results demonstrate that the ribosome–oligonucleotide interactions that anchor the bead onto the ribosome are sufficiently strong for force measurements within the required range. In addition, post-rupture colocalization of the bead and the ribosome complex with fluorescent tetramethylrhodamine-labelled[11] tRNA in the P site suggests that the rupture involves disruption of tRNA–mRNA interactions, because after the rupture event the mRNA is attached to the surface while the tRNA remains on the bead-associated ribosome (data not shown). This provided further proof that the observed rupture does not occur at the ribosome–oligonucleotide–bead linkage.

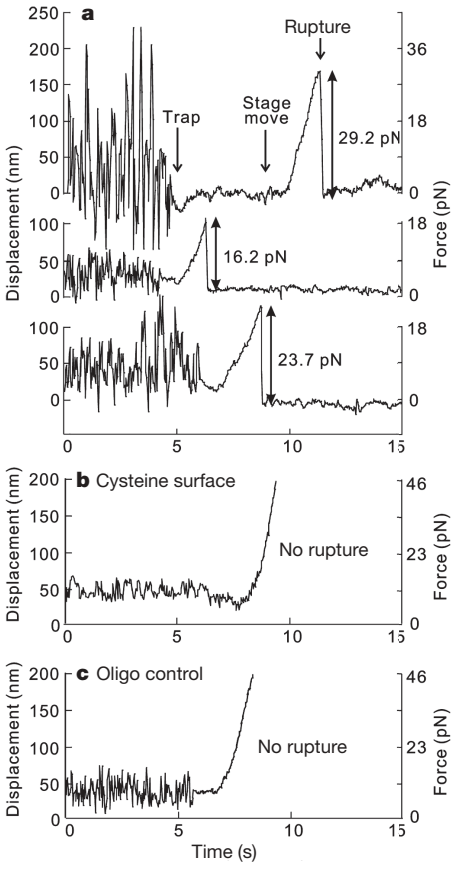


Figure 2 | Examples of displacement–time traces, showing the behaviour of the bead. **a**, The position of tethered beads fluctuates around the tether centre. Once the bead is trapped by the optical tweezers, its fluctuations become suppressed as indicated. As the bead starts to follow the stage movement, the force exerted on the complex increases. When the ribosome complex ruptures, the bead returns to the trapping centre. **b**, Control measurement for ribosomes covalently crosslinked to a cysteine-reactive surface. No rupture events were observed within our measurement range. **c**, Control measurement for a biotinylated RNA oligonucleotide designed to mimic the extension in the 16S rRNA. The RNA oligonucleotide was attached to a streptavidin-derivatized surface and hybridized with the DNA oligonucleotide–bead conjugate. No rupture events were observed.

Binding of tRNAs to the ribosome stabilizes ribosome–mRNA interactions. The rupture force distribution for a ribosome–mRNA complex in the absence of tRNAs showed a single peak at 10.6 pN (Fig. 3a). The addition of non-acylated initiator tRNA^{fMet} to the P site strengthened the ribosome–mRNA interactions roughly by 5 pN, increasing the rupture force to 15.2 pN (Fig. 3b). Subsequent addition of Phe-tRNA^{Phe} to the A site resulted in further stabilization of the complex by another 10 pN, giving a single peak distribution centred at 26.5 pN (Fig. 3c). In all experiments, the tRNA occupancy of each ribosome complex was verified just before the force measurement using fluorescence resonance energy transfer (FRET). For this purpose, tRNA^{fMet} in the P site was labelled with Cy3, Phe-tRNA^{Phe} in the A site was labelled with Cy5, and fluorescence of both dyes was monitored at 532 nm excitation following methods described previously[12] (Supplementary Fig. 1). Only complexes showing FRET owing to the presence of both tRNAs were included in further force measurement analysis, so that the final force population distributions are representative of ribosomal complexes with both A and P sites occupied (Fig. 3c–f, i, j). The multiple interactions between P-site and A-site tRNAs and both the large and small ribosomal subunits[6,7] probably contribute to the stabilization of the ribosomal complex on the mRNA and explain the large rupture force increase upon the addition of tRNAs.

The nature of the aminoacyl group on A-site tRNA strongly affects ribosome–mRNA rupture forces. Binding of the peptidyl-tRNA analogue N-acetyl-Phe-tRNA^{Phe} to the A site, which mimics the post-peptidyl transfer state, resulted in a significant reduction of the rupture force to 12.7 pN (Fig. 3d), suggesting that the interactions between the ribosome and mRNA are weakened after peptide bond formation. The lower affinity of the peptidyl-tRNA in the A site[13] might also contribute to the decrease in rupture force. Complexes prepared with fMet-tRNA^{fMet} in the P site and Phe-tRNA^{Phe} in the A site, allowing for efficient formation of a ribosome-catalysed peptide bond, showed a weak rupture force centred at 11.4 pN, with only a minor peak at 24.8 pN most probably representative of a small subpopulation (14.3%) of either inactive ribosomes and/or ribosomes where the P-site tRNA has become deacylated (Fig. 3e). This provides further evidence that the presence of the peptidyl-tRNA moiety in the A site results in the destabilization of ribosome–mRNA interactions. In addition, the stability of the ribosome on mRNA is not affected by the equilibrium between the classical and hybrid tRNA states. In related work, we have observed that the post-peptide bond mimic (tRNA^{fMet} in the P site and N-acetyl-Phe-tRNA^{Phe} in the A site) results in rapid classical–hybrid state fluctuations at 5 mM Mg^{2+} (with the relative occupancy of classical and hybrid states of 3:2), while the tRNAs remain predominantly in the classical configuration at 15 mM Mg^{2+} (H.D.K. *et al.*, manuscript in preparation). No significant difference was observed in the rupture force distributions of the post-peptidyl transfer complex in 5 mM Mg^{2+} and 15 mM Mg^{2+} buffers (compare Fig. 3d and f). Whereas different Mg^{2+} concentrations can affect tRNA–ribosome complex stability[13], the timescale of

tRNA dissociation is much greater than the time it took to complete the force measurements. The observed lack of strong Mg^{2+} dependence to the measured forces indicates that we are monitoring rupture of mRNA–ribosome base pairings, which are not strongly Mg^{2+} dependent.

The SD interaction directly increases the binding affinity of the ribosome for mRNA and influences tRNA–mRNA translocation[5,14]. To test the contribution of the SD interaction to complex stability, rupture measurements were performed on ribosomal complexes assembled on an mRNA where the SD interaction had been significantly weakened by modifying the sequence from AGGA to ACCA[14]. In the absence of tRNAs, no tethered beads were observed, indicating that the complexes without the SD region were too unstable for a force measurement (Fig. 3g). The addition of $tRNA^{fMet}$ to the P site stabilized the complexes by almost 5 pN, which is similar to the stabilization observed for the SD-containing complexes, allowing for a rupture force measurement of 4.8 pN (Fig. 3h). Further addition of Phe-$tRNA^{Phe}$ to the A site increased the rupture force to 14.8 pN (Fig. 3i), a 10 pN increase again similar to that observed for the SD-containing complexes. Upon binding of N-acetyl-Phe-$tRNA^{Phe}$ to the complexes without the SD sequence, a rupture force of 12.1 pN was measured (Fig. 3j), slightly reduced in comparison to the force observed for Phe-$tRNA^{Phe}$ complexes (Fig. 3i). This final post-peptidyl transfer rupture force was essentially the same for the SD-containing and SD-lacking complexes (compare Fig. 3d and j), which raises the possibility that peptide bond formation results in destabilization of SD interactions between the mRNA and the small ribosomal subunit.

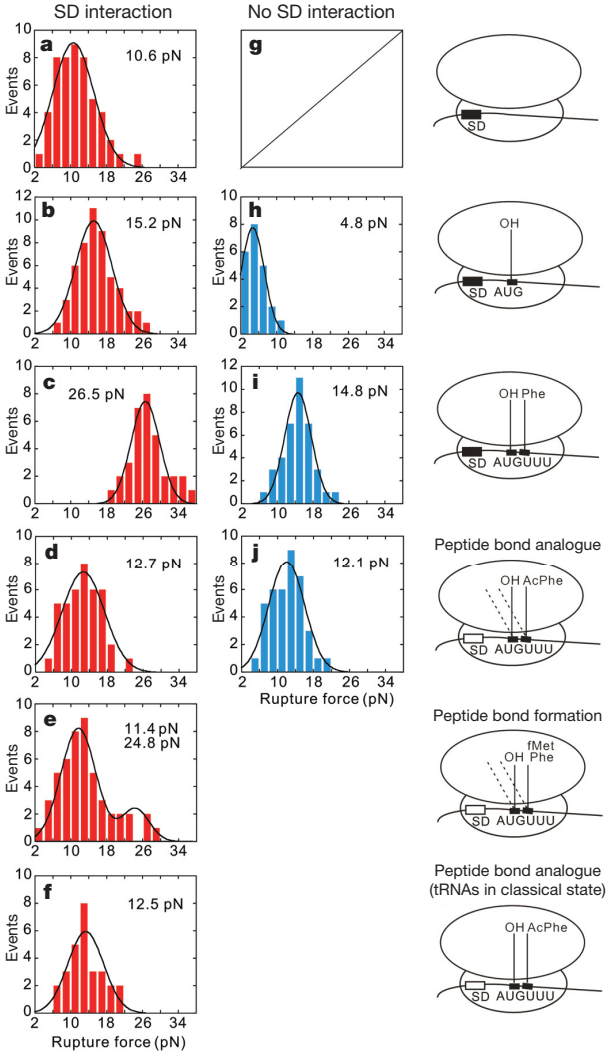


Figure 3 | Rupture force distributions for ribosome complexes assembled on mRNAs containing (left panels) or lacking (middle panels) the SD sequence. The numbers above each histogram are the force values determined for each complex. All complexes were assembled in 5 mM Mg^{2+} unless indicated otherwise. **a**, **g**, Ribosome–mRNA complex without tRNAs. No tethered beads were observed in the absence of the SD sequence (**g**), indicating that the complex was too unstable for a force measurement under these conditions. **b**, **h**, Ribosome–mRNA complex carrying deacylated $tRNA^{fMet}$ in the P site. **c**, **i**, Ribosome–mRNA complex with $tRNA^{fMet}$ in the P site and Phe-$tRNA^{Phe}$ in the A site. **d**, **j**, Ribosome–mRNA complex with $tRNA^{fMet}$ in the P site and N-acetyl-Phe-$tRNA^{Phe}$ in the A site. **e**, Ribosome–mRNA complex after ribosome-catalysed peptide bond formation. The complex was assembled with fMet-$tRNA^{fMet}$ in the P site and Phe-$tRNA^{Phe}$ in the A site, and incubated for 20 min to allow for peptidyl transfer. **f**, Ribosome–mRNA complex with $tRNA^{fMet}$ in the P site and N-acetyl-Phe-$tRNA^{Phe}$ in the A site in 15 mM Mg^{2+}. The diagrams on the right show the composition of the complexes containing the SD sequence: strong SD interactions are indicated by black boxes, while destabilized SD interactions are shown as white boxes.

Although the exact rupture pathway cannot be determined for the different ribosomal complexes, Cy3 and Cy5 fluorescence analysis verified the initial tRNA occupancy on the ribosomes. Owing to the rapid photobleaching of Cy3 in the presence of high intensity optical trapping light, we could not confirm tRNA occupancy after the force measurement using the same dye pair. Separate experiments using tetramethylrhodamine-labelled tRNA to demonstrate post-rupture colocalization of the ribosome complex and the bead suggest that the rupture occurs through the disruption of interactions between the 30S–tRNA complexes and the mRNA, as the 30S subunit and at least one bound tRNA remain attached to the bead, while the mRNA stays anchored on the surface via the strong biotin–streptavidin link. As both the initial and the final ligand-bound states are comparable for the different ribosome complexes, the rupture pathways are presumed to be similar and are unlikely to involve an intermediate spontaneous translocation step (known to occur in the absence of EF-G at a very slow rate[15]) or spontaneous dissociation of the entire ribosomal complex on the timescale of the force measurement experiment.

Our results provide direct evidence for coupling of the 50S peptidyl transferase centre and the 30S subunit. Precisely how the formation of a peptide bond on the 50S portion of the ribosome leads to the weakening of the mRNA contacts with the 30S subunit needed for processive translation is yet to be determined. Previous studies have demonstrated that the ribosome can sense the chemical nature of its ligand in the A site: peptidyl-tRNA shows much lower binding affinity than aminoacyl-tRNA[13], while deacylation of the A-site bound tRNA can affect the accuracy of translocation along mRNA[14]. The distinct interactions of aminoacyl-tRNA versus peptidyl-tRNA moieties at the peptidyl transferase centre after peptide bond formation might induce conformational changes within the 50S subunit, which could be further propagated through the subunit interface towards the anti-SD region of the 30S subunit, possibly via a relative movement of the two subunits[16], resulting in the destabilization of SD interactions between the 30S subunit and mRNA, and facilitating the first translocation step. A mechanistic explanation of this allosteric communication requires further investigation.

Received 29 July 2006; accepted 26 January 2007.

1. Noller, H. F. Structure of ribosomal RNA. *Annu. Rev. Biochem.* **53**, 119–162 (1984).
2. Wintermeyer, W. *et al.* Mechanisms of elongation on the ribosome: dynamics of a macromolecular machine. *Biochem. Soc. Trans.* **32**, 733–737 (2004).
3. Green, R. & Noller, H. F. Ribosomes and translation. *Annu. Rev. Biochem.* **66**, 679–716 (1997).
4. Shine, J. & Dalgarno, L. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl Acad. Sci. USA* **71**, 1342–1346 (1974).

5. Calogero, R. A., Pon, C. L., Canonaco, M. A. & Gualerzi, C. O. Selection of the mRNA translation initiation region by *Escherichia coli* ribosomes. *Proc. Natl Acad. Sci. USA* **85,** 6427–6431 (1988).
6. Yusupov, M. M. *et al.* Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292,** 883–896 (2001).
7. Yusupova, G. Z., Yusupov, M. M., Cate, J. H. D. & Noller, H. F. The path of messenger RNA through the ribosome. *Cell* **106,** 233–241 (2001).
8. Moazed, D. & Noller, H. F. Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell* **57,** 585–597 (1989).
9. Moazed, D. & Noller, H. F. Binding of tRNA to the ribosomal A and P sites protects two distinct sets of nucleotides in 16S rRNA. *J. Mol. Biol.* **211,** 135–145 (1990).
10. Dorywalska, M. *et al.* Site-specific labeling of the ribosome for single-molecule spectroscopy. *Nucleic Acids Res.* **33,** 182–189 (2005).
11. van Dijk, M. A., Kapitein, L. C., van Mameren, J., Schmidt, C. F. & Peterman, E. J. G. Combining optical trapping and single-molecule fluorescence spectroscopy: Enhanced photobleaching of fluorophores. *J. Phys. Chem. B* **108,** 6479–6484 (2004).
12. Blanchard, S. C., Kim, H. D., Gonzalez, R. L. Jr, Puglisi, J. D. & Chu, S. tRNA dynamics on the ribosome during translation. *Proc. Natl Acad. Sci. USA* **101,** 12893–12898 (2004).
13. Semenkov, Y. P., Rodnina, M. V. & Wintermeyer, W. Energetic contribution of tRNA hybrid state formation to translocation catalysis on the ribosome. *Nature Struct. Biol.* **7,** 1027–1031 (2000).
14. Fredrick, K. & Noller, H. F. Accurate translocation of mRNA by the ribosome requires a peptidyl group or its analog on the tRNA moving into the 30S P site. *Mol. Cell* **9,** 1125–1131 (2002).
15. Katunin, V. I., Savelsbergh, A., Rodnina, M. V. & Wintermeyer, W. Coupling of GTP hydrolysis by elongation factor G to translocation and factor recycling on the ribosome. *Biochemistry (Mosc.)* **41,** 12806–12812 (2002).
16. Frank, J. & Agrawal, R. K. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* **406,** 318–322 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Squires for the SQ380 strain that was used to engineer the ribosomes. We are grateful to colleagues in the S. Chu and J. Puglisi groups at Stanford University for encouragement and discussions. S.U. is a recipient of JSPS Postdoctoral Fellowships for Research Abroad. M.D. was supported by the HHMI Predoctoral Fellowship. This work was funded by grants to J.D.P. from the NIH and the Packard Foundation, to S.C. from the NSF and NASA, and to J.D.P. and S.C. from the Packard Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.D.P. (puglisi@stanford.edu) or S.C. (schu@lbl.gov).

LETTERS

Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein

Michael D. Stone[1], Mariana Mihalusova[2], Catherine M. O'Connor[5], Ramadevi Prathapam[5], Kathleen Collins[5] & Xiaowei Zhuang[1,3,4]

Telomerase is an essential cellular ribonucleoprotein (RNP) that solves the end replication problem and maintains chromosome stability by adding telomeric DNA to the termini of linear chromosomes[1–3]. Genetic mutations that abrogate the normal assembly of telomerase RNP cause human disease[4]. It is therefore of fundamental and medical importance to decipher cellular strategies for telomerase biogenesis, which will require new insights into how specific interactions occur in a precise order along the RNP assembly pathway. Here we use a single-molecule approach to dissect the individual assembly steps of telomerase. Direct observation of complex formation in real time revealed two sequential steps of protein-induced RNA folding, establishing a hierarchical RNP assembly mechanism: interaction with the telomerase holoenzyme protein p65 induces structural rearrangement of telomerase RNA, which in turn directs the binding of the telomerase reverse transcriptase to form the functional ternary complex. This hierarchical assembly process is facilitated by an evolutionarily conserved structural motif within the RNA. These results identify the RNA folding pathway during telomerase biogenesis and define the mechanism of action for an essential telomerase holoenzyme protein.

Telomerase RNP functions as a multisubunit holoenzyme consisting of telomerase RNA, telomerase reverse transcriptase (TERT) and additional protein cofactors. Catalytically active telomerase enzyme can be reconstituted from RNA and TERT in rabbit reticulocyte lysate (RRL), in which general chaperone activities promote RNP assembly[5,6]. However, the endogenous process of telomerase biogenesis seems to require a more specific assembly pathway[7]. In support of this view, cellular accumulation of telomerase RNP is promoted by a number of specific RNA-binding proteins, including dyskerin in vertebrates, Sm proteins in yeasts, and La-motif proteins in ciliates[8–11]. In the present study we exploited single-molecule fluorescence resonance energy transfer (FRET)[12–14] to explore the mechanism for telomerase RNP biogenesis, using the ciliate *Tetrahymena thermophila* as a model system.

The *Tetrahymena* telomerase RNA is a 159-nucleotide transcript (Fig. 1a) that provides a template for telomere synthesis and functions in adapting the polymerase to its specialized task of reiterative repeat synthesis[15–17]. Using a refined DNA-splinted RNA ligation method[18] (Supplementary Fig. 1) we strategically placed a FRET donor (Cy3) and acceptor (Cy5) flanking the regions important for interaction with TERT and the holoenzyme protein p65, a La-motif protein that promotes telomerase RNP accumulation *in vivo*[11] (Fig. 1a). To facilitate real-time observation of telomerase RNP assembly, RNA was surface-immobilized through an extension of stem II that does not perturb telomerase activity *in vitro* or *in vivo*[19]. Standard DNA primer extension assays performed with both labelled and unlabelled RNA yielded normal profiles of product synthesis (Fig. 1a, inset), indicating that dye labelling did not substantially interfere with RNA function.

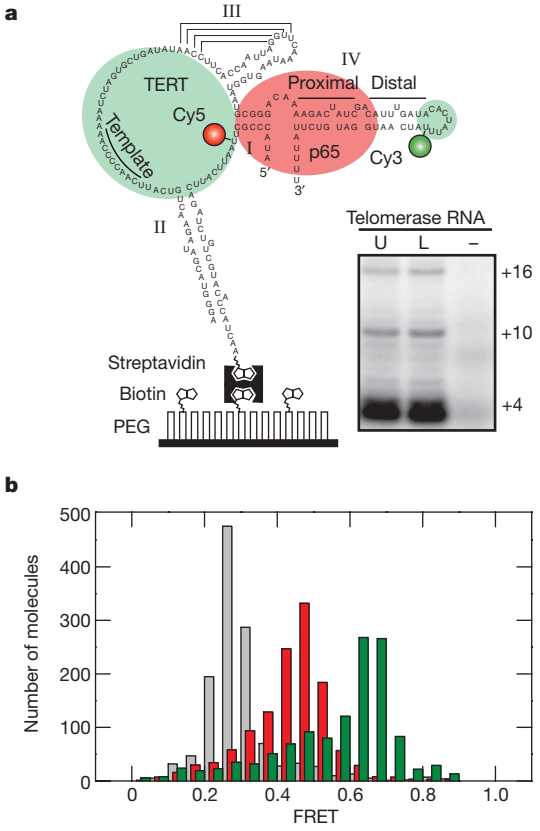


Figure 1 | Telomerase proteins p65 and TERT induce distinct conformational changes in telomerase RNA. a, Full-length telomerase RNA labelled with FRET donor (Cy3) and acceptor (Cy5). RNA molecules were immobilized on a streptavidin-coated surface by a biotin molecule engineered onto an extension of stem II. The interaction sites with p65 (red) and TERT (green) are highlighted[20,24]. Inset: telomerase primer extension assay with dye-labelled (L) and unlabelled (U) RNA or without RNA (−). Numbers at the right indicate the number of nucleotides added to generate each product. **b**, FRET histograms of RNA molecules in the absence of protein (grey bars), the presence of 10 nM p65 (red bars) or 10 nM p65 plus 32 nM $TERT_{1–516}$ (green bars). A peak at FRET = 0 due to the presence of about 30% of molecules without active Cy5 (Supplementary Fig. 2) was removed from these histograms.

[1]Department of Chemistry and Chemical Biology, [2]Department of Molecular and Cellular Biology, [3]Department of Physics, and [4]Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA. [5]Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA.

The FRET distribution for full-length telomerase RNA showed a single population centred at FRET = 0.29 (Fig. 1b and Supplementary Fig. 2a). Time traces of individual RNA molecules showed a stable FRET level without significant fluctuations except for transient excursions to FRET = 0 due to blinking of the acceptor dye (Supplementary Fig. 3). Addition of purified p65 gave rise to a second population with a steady FRET level at 0.46 (Fig. 1b and Supplementary Fig. 3). Estimating the relative abundance of each population as a function of p65 concentration yielded a non-cooperative p65 binding isotherm with $K = 1.3$ nM (Supplementary Fig. 4), in agreement with ensemble characterization of p65–RNA affinity[20]. Similar results were obtained with telomerase RNA immobilized by a 5′ extension (Supplementary Fig. 5), showing that surface attachment did not perturb the assembly of telomerase RNP. Control experiments designed to prevent a distance change between donor and acceptor dyes upon binding of p65 showed that direct dye–p65 interaction, if present, did not induce any significant change in FRET (Supplementary Fig. 6). Furthermore, binding of p65 did not alter fluorescence intensities from the RNA singly labelled with Cy3 or Cy5 at the same locations (data not shown). Taken together, these results indicate that the p65-induced increase in FRET represents a conformational change within the RNA, wherein the ends of stem I and stem IV are brought closer together in the tertiary RNA fold.

To investigate the assembly of p65–RNA–TERT ternary complex, we used a purified TERT polypeptide containing the amino-terminal 516 amino acids ($TERT_{1–516}$). This polypeptide contains all of the primary sites of RNA–TERT interaction, but unlike the full-length TERT it can be expressed in a soluble recombinant form[20]. Addition of $TERT_{1–516}$ to the p65–RNA complex further shifted the FRET value to a steady value of 0.65 (Fig. 1b and Supplementary Fig. 3), with a binding affinity of $K = 14.4$ nM (Supplementary Fig. 4). The

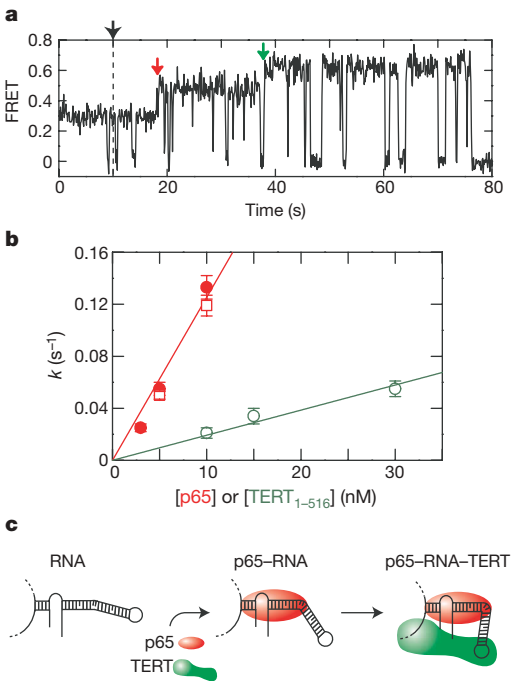


Figure 2 | Real-time assembly of individual telomerase RNP complexes. **a**, A single-molecule FRET trajectory showing hierarchical RNP assembly after the addition (black arrow) of a protein mixture of p65 (10 nM) and $TERT_{1–516}$ (10 nM), characterized by a p65-induced FRET change (red arrow) followed by a second FRET transition after complete assembly of the ternary p65–RNA–$TERT_{1–516}$ complex (green arrow). Excursions to FRET = 0 correspond to blinking of Cy5 (Supplementary Fig. 3) **b**, The conversion rate (k) from FRET = 0.29 to FRET = 0.46 (red) or from FRET = 0.46 to FRET = 0.65 (green) as a function of p65 or $TERT_{1–516}$ concentration, respectively. Open red squares indicate conversion rates measured in the presence of p65 only, and solid red circles indicate those measured in the presence of both p65 and $TERT_{1–516}$. Results are means ± s.e.m. ($n \geq 32$ for all conditions). Linear fits of the data (red and green lines) yield binding rate constants of 1.3×10^7 and 2×10^6 M^{-1} s^{-1} for p65 and $TERT_{1–516}$, respectively. **c**, The interaction of p65 (red) induces compaction in stem I–IV of the RNA, followed by co-assembly with TERT and concomitant additional RNA folding in the ternary complex. The conserved GA bulge introduces a kink in stem IV in the absence of protein[25,26]. The additional bending in stem IV in the presence of proteins is drawn to indicate RNA compaction as measured by FRET, and is not meant to represent a specific structural orientation. Note that tertiary RNA contacts not shown in the model may be formed within the telomerase holoenzyme.

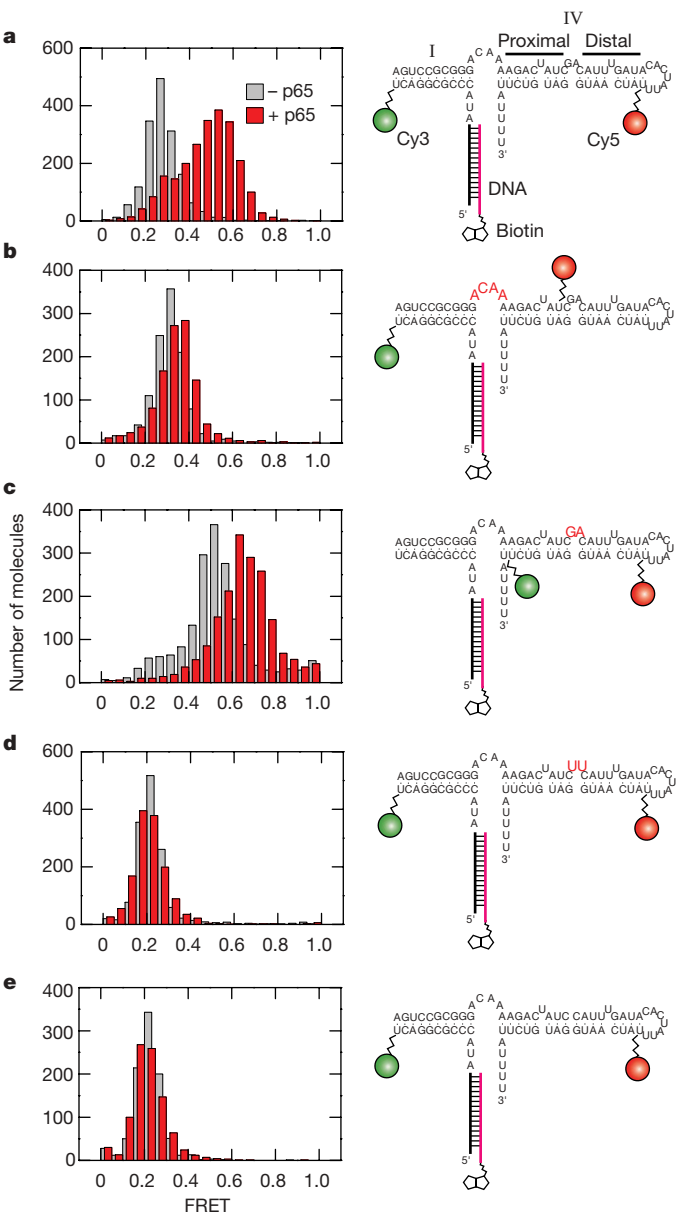


Figure 3 | The p65-induced RNA structural change occurs within stem IV and requires the conserved GA bulge. Schematic illustrations of stem I–IV constructs (right panels) and their corresponding FRET histograms (left panels) in the absence (grey) or presence (red) of p65. **a–c**, Constructs probing conformational change within the entire stem I and stem IV region (**a**), between stem I and proximal stem IV (**b**) and between proximal stem IV and distal stem IV (**c**). **d**, **e**, Constructs probing the role of the GA bulge by substitution (GA→UU) (**d**) and by deletion (ΔGA) (**e**). Stem I–IV constructs were labelled with FRET donor (green spheres) and acceptor (red spheres) and surface-immobilized by means of a 5′ RNA extension (thick black line) annealed to a biotinylated complementary DNA oligonucleotide (pink line).

binding of $TERT_{1–516}$ did not alter fluorescence intensities from RNA singly labelled with Cy3 or Cy5 (data not shown), indicating that the $TERT_{1–516}$-induced FRET change signifies another genuine structural rearrangement in telomerase RNA. FRET measurements on catalytically active telomerase RNPs assembled in RRL with p65 and full-length TERT gave a similar FRET distribution (compare Supplementary Fig. 7a with Fig. 1b), showing that the high FRET conformation (FRET = 0.65) observed in the p65–RNA–$TERT_{1–516}$ ternary complex represents the RNA structure in the functional enzyme.

The distinct FRET signatures of the telomerase RNA alone, the p65–RNA complex and the p65–RNA–$TERT_{1–516}$ ternary complex allowed us to examine the assembly of individual telomerase RNP particles in real time. We therefore added protein mixtures containing both p65 and $TERT_{1–516}$ to telomerase RNA and monitored the FRET signal of individual RNA molecules during the assembly process. Most assembly trajectories (74%, $n = 173$) displayed a prominent kinetic intermediate with FRET = 0.46, equivalent to that measured from p65–RNA complex, before attaining a final state with FRET = 0.65 indicative of the p65–RNA–TERT ternary complex (Fig. 2a and Supplementary Fig. 3b). The remaining minor fraction of assembly trajectories exhibited less defined intermediates with irregular FRET fluctuations (Supplementary Fig. 8). The fraction of molecules assembled through the well-defined two-step (0.29→0.46→0.65) pathway did not vary substantially with altered ratios of p65 to $TERT_{1–516}$ concentrations (at p65-to-$TERT_{1–516}$ ratios of 1:1, 1:3 and 1:10, the proportions were 73% ($n = 37$), 76% ($n = 101$) and 69% ($n = 35$), respectively). The rates of conversion from FRET = 0.29 to FRET = 0.46 and from FRET = 0.46 to FRET = 0.65 increased linearly with p65 and $TERT_{1–516}$ concentrations, respectively (Fig. 2b).These results indicate a possible hierarchical telomerase assembly process directed by sequential steps of protein-induced RNA folding: assembly is initiated by p65 binding, stabilizing a RNA structural intermediate, which in turn promotes the functional co-assembly of TERT with telomerase RNA (Fig. 2c).

To characterize the p65-induced assembly intermediate structurally, we generated a series of truncated constructs composed of telomerase RNA stems I and IV, harbouring the primary p65 interaction sites[20] (Fig. 3). When FRET donor and acceptor were incorporated in stem I and distal stem IV (approximating the labelling sites in the full-length RNA construct), we observed a significant increase in FRET when p65 bound (Fig. 3a), which is consistent with results described above (Fig. 1b). Experiments with FRET donor and acceptor relocated to flank the ACAA linker (Fig. 3b) or the dinucleotide GA bulge (Fig. 3c) indicated that the primary p65-induced structural rearrangement occurs between proximal and distal stem IV linked by the GA bulge rather than between stem I and stem IV. Furthermore, substitution (GA→UU; Fig. 3d) or deletion (ΔGA; Fig. 3e) of the GA bulge abolished any stable p65-induced FRET change. At the concentration of p65 used here (30 nM), most RNA molecules were bound to p65 (Supplementary Fig. 9). Thus, the p65-induced RNA conformational change occurs within stem IV and requires the central stem IV GA bulge, which is conserved across all *Tetrahymena* species[21].

To probe whether the p65-induced RNA conformation is an essential intermediate during hierarchical telomerase RNP biogenesis, we compared the assembly observations described above with reactions that lacked p65 or lacked the GA bulge in telomerase RNA. Telomerase RNA in the presence of $TERT_{1–516}$ alone did not yield the stable high-FRET state (FRET = 0.65) observed for the p65–RNA–$TERT_{1–516}$ ternary complex (compare Fig. 4a with Fig. 1b). Instead, the FRET trajectories showed characteristic fluctuations with amplitudes and rates that were independent of $TERT_{1–516}$ concentration (1–100 nM) (Fig. 4b and Supplementary Fig. 10), indicating that the FRET dynamics were not due to the repetitive binding and dissociation of protein. The fraction of molecules showing such fluctuations increased with the concentration of $TERT_{1–516}$ (data not shown), with an affinity comparable to the previously determined value for telomerase RNA and $TERT_{1–516}$ (ref. 20). These results show that the RNA fold in the RNA–$TERT_{1–516}$ complex is structurally different from that in the p65–RNA–$TERT_{1–516}$ ternary complex. Addition of p65 to preformed RNA–$TERT_{1–516}$ complexes recovered the stable high-FRET conformation (data not shown), indicating that p65 might be able to rescue misassembled RNA–TERT complexes by directing productive interactions between TERT and telomerase RNA. Furthermore, p65 substantially stimulated the complex formation and catalytic activity of telomerase RNPs reconstituted in RRL (Supplementary Fig. 7). The role of p65 therefore cannot be fully recapitulated by heterologous factors within RRL.

When p65 was added to full-length telomerase RNA lacking the GA bulge, no stable FRET change was observed (Fig. 4c). Instead, the FRET traces visited a higher FRET state only transiently (Fig. 4d), with a rate independent of p65 concentration (data not shown), indicating that disruption of the GA bulge might render the stem IV region too rigid to be folded by p65 in a stable manner. Subsequent addition of $TERT_{1–516}$ did not substantially change the FRET dynamics or induce the FRET = 0.65 conformation indicative of the functional p65–RNA–TERT complex (compare Fig. 4c with

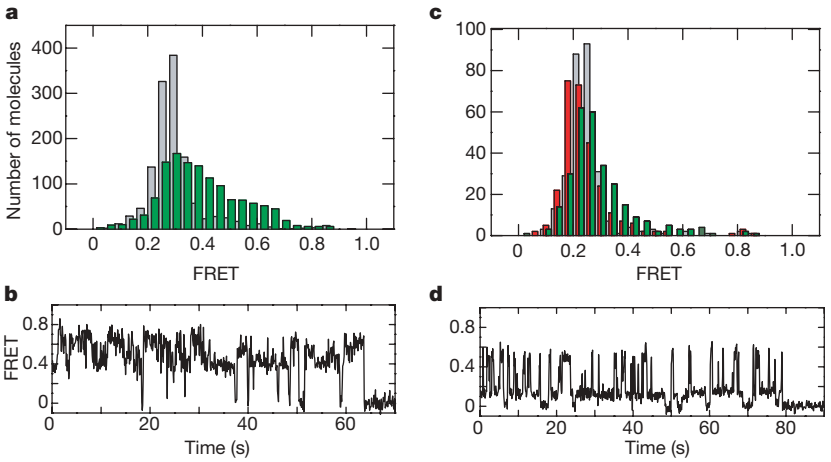


Figure 4 | The p65 protein and conserved GA bulge in stem IV of telomerase RNA are required for telomerase assembly. a, FRET histograms of full-length RNA in the absence (grey bars) or presence (green bars) of 10 nM $TERT_{1–516}$. **b**, A single-molecule FRET trajectory of telomerase RNA in the presence of 10 nM $TERT_{1–516}$. **c**, FRET histograms of full-length RNA lacking the GA bulge (ΔGA) in the absence of protein (grey bars), in the presence of 10 nM p65 (red bars) or in the presence of 10 nM p65 plus 30 nM $TERT_{1–516}$ (green bars). **d**, A single-molecule FRET trajectory of ΔGA RNA in the presence of 10 nM p65.

Fig. 1b). Furthermore, *in vivo* deletion of the conserved GA bulge or its replacement with UU prevented the incorporation of telomerase RNA into a stable RNP, resulting in the failure of these RNA variants to accumulate relative to endogenous telomerase RNA (Supplementary Fig. 11). Taken together, these results show that the p65-induced RNA conformational change is critical for formation of the native p65–RNA–TERT complex.

Thus, our experiments directly show a hierarchical assembly mechanism for telomerase RNP in which the protein subunits mould a specific RNA tertiary structure in a stepwise fashion. First, the telomerase holoenzyme protein p65 induces a marked structural change within the stem IV region of the RNA, a domain important for processive telomere synthesis[15–17]. This stable structural rearrangement strictly requires the central stem IV GA bulge, accounting for the phylogenetic conservation of this sequence motif. Second, the RNA conformation in the p65–RNA complex is further altered by binding of TERT, resulting in a compact RNA tertiary fold within the functional telomerase RNP. By monitoring the assembly of individual RNP complexes in real time, we showed that the p65-induced RNA conformation is a fundamental structural intermediate during hierarchical telomerase assembly. These results provide a likely mechanism for telomerase RNP assembly *in vivo*, in which a holoenzyme protein repositions distant RNA-binding sites for TERT, promoting the correct co-assembly of the catalytic subunit TERT with telomerase RNA (Figs 1a and 2c). Consistent with this notion, removal of p65 (ref. 11) or mutation of the GA bulge prevents the incorporation of telomerase RNA into a stable RNP *in vivo* (Supplementary Fig. 11). An analogous hierarchical biogenesis pathway has been proposed for the ribosome[22,23], which contains an RNA rather than a protein catalytic centre. Hierarchical steps of protein-mediated RNA folding might therefore represent a general assembly strategy for cellular RNPs, providing multiple layers of control to circumvent the formation of long-lived misassembled intermediates.

METHODS

RNA and protein preparation. Full-length FRET-labelled RNAs were prepared by a refined DNA-splinted RNA ligation method detailed in Supplementary Methods and Supplementary Fig. 1. The preparation of truncated RNA constructs is described in Supplementary Methods. Histidine-tagged p65 and $TERT_{1–516}$ proteins were expressed and purified from *E. coli* as described[20].

Single-molecule FRET. Single-molecule FRET measurements were performed as described in Supplementary Methods. The FRET value is defined as $I_A/(I_A + I_D)$, where I_D and I_A are the fluorescence intensities measured in the donor and acceptor channels, respectively.

RRL reconstitution of telomerase RNP and *in vivo* RNA accumulation assays. Protein expression in RRL was performed by standard methods (Supplementary Methods) with the pCITE plasmid system as described[16]. *In vivo* telomerase RNA accumulation assays are described in Supplementary Methods.

Received 18 September 2006; accepted 15 January 2007.
Published online 25 February 2007.

1. Greider, C. W. & Blackburn, E. H. Tracking telomerase. *Cell* **116,** S83–S86 (2004).
2. Cech, T. R. Beginning to understand the end of the chromosome. *Cell* **116,** 273–279 (2004).
3. Collins, K. The biogenesis and regulation of telomerase holoenzymes. *Nature Rev. Mol. Cell Biol.* **7,** 484–494 (2006).
4. Wong, J. M. & Collins, K. Telomere maintenance and disease. *Lancet* **362,** 983–988 (2003).
5. Weinrich, S. L. *et al.* Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTRT. *Nature Genet.* **17,** 498–502 (1997).
6. Holt, S. E. *et al.* Functional requirement of p23 and Hsp90 in telomerase complexes. *Genes Dev.* **13,** 817–826 (1999).
7. Harrington, L. Biochemical aspects of telomerase function. *Cancer Lett.* **194,** 139–154 (2003).
8. Mitchell, J. R., Wood, E. & Collins, K. A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* **402,** 551–555 (1999).
9. Seto, A. G., Zaug, A. J., Sobel, S. G., Wolin, S. L. & Cech, T. R. *Saccharomyces cerevisiae* telomerase is an Sm small nuclear ribonucleoprotein particle. *Nature* **401,** 177–180 (1999).
10. Aigner, S., Postberg, J., Lipps, H. J. & Cech, T. R. The *Euplotes* La motif protein p43 has properties of a telomerase-specific subunit. *Biochemistry* **42,** 5736–5747 (2003).
11. Witkin, K. L. & Collins, K. Holoenzyme proteins required for the physiological assembly and activity of telomerase. *Genes Dev.* **18,** 1107–1118 (2004).
12. Stryer, L. & Haugland, R. P. Energy transfer: a spectroscopic ruler. *Proc. Natl Acad. Sci. USA* **58,** 719–726 (1967).
13. Ha, T. *et al.* Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl Acad. Sci. USA* **93,** 6264–6268 (1996).
14. Zhuang, X. Single-molecule RNA science. *Annu. Rev. Biophys. Biomol. Struct.* **34,** 399–414 (2005).
15. Sperger, J. M. & Cech, T. R. A stem-loop of *Tetrahymena* telomerase RNA distant from the template potentiates RNA folding and telomerase activity. *Biochemistry* **40,** 7005–7016 (2001).
16. Lai, C. K., Miller, M. C. & Collins, K. Roles for RNA in telomerase nucleotide and repeat addition processivity. *Mol. Cell* **11,** 1673–1683 (2003).
17. Mason, D. X., Goneska, E. & Greider, C. W. Stem-loop IV of *Tetrahymena* telomerase RNA stimulates processivity in *trans*. *Mol. Cell. Biol.* **23,** 5606–5613 (2003).
18. Moore, M. J. & Query, C. C. Joining of RNAs by splinted ligation. *Methods Enzymol.* **317,** 109–123 (2000).
19. Cunningham, D. D. & Collins, K. Biological and biochemical functions of RNA in the *Tetrahymena* telomerase holoenzyme. *Mol. Cell. Biol.* **25,** 4442–4454 (2005).
20. Prathapam, R., Witkin, K. L., O'Connor, C. M. & Collins, K. A telomerase holoenzyme protein enhances telomerase RNA assembly with telomerase reverse transcriptase. *Nat. Struct. Mol. Biol.* **12,** 252–257 (2005).
21. Ye, A. J. & Romero, D. P. Phylogenetic relationships amongst tetrahymenine ciliates inferred by a comparison of telomerase RNAs. *Int. J. Syst. Evol. Microbiol.* **52,** 2297–2302 (2002).
22. Adilakshmi, T., Ramaswamy, P. & Woodson, S. A. Protein-independent folding pathway of the 16S rRNA 5′ domain. *J. Mol. Biol.* **351,** 508–519 (2005).
23. Talkington, M. W., Siuzdak, G. & Williamson, J. R. An assembly landscape for the 30S ribosomal subunit. *Nature* **438,** 628–632 (2005).
24. O'Connor, C. M., Lai, C. K. & Collins, K. Two purified domains of telomerase reverse transcriptase reconstitute sequence-specific interactions with RNA. *J. Biol. Chem.* **280,** 17533–17539 (2005).
25. Richards, R. J. *et al.* Structural study of elements of *Tetrahymena* telomerase RNA stem–loop IV domain important for function. *RNA* **12,** 1475–1485 (2006).
26. Chen, Y. *et al.* Structure of stem-loop IV of *Tetrahymena* telomerase RNA. *EMBO J.* **25,** 3156–3166 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Bates for LabView software for data acquisition. This work was supported in part by the NIH and the Packard Foundation (X.Z.), and the NIH (K.C.). X.Z. is a Howard Hughes Medical Institute investigator. M.D.S. is a. NIH Ruth L. Kirschstein NSRA Fellow.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to X.Z. (zhuang@chemistry.harvard.edu).

naturejobs

THE CAREERS MAGAZINE FOR SCIENTISTS

It's strange, but recently all I seem to hear about are young scientists whose relationships have broken down. OK, that's a slight exaggeration. But I did run into one of *Naturejobs*' former writers for Graduate Journal at a postdoc retreat the other week, only to find out that they had just undergone a divorce. Not long after, a chat with a friend revealed that her brother's relationship had ended when he began his PhD studies. These tales of woe reminded me of another Graduate Journal writer who got divorced while he was working on his dissertation. And that got me thinking about the effect that building a scientific career can have on relationships.

I don't have any facts or figures about relationship issues among PhD scientists, but I suspect that they get married later than the average, and experience more romantic difficulties than normal earlier in their career. But why? First, I suspect that location can exert a significant pressure. When someone selects a postdoc or a first job, it's unlikely to be close to wherever a spouse or a partner already works — especially if the partner is also in academia. That means young scientists must either limit their job search, expect their partner to move with them, or carry out a long-distance relationship. A choice that is less than satisfactory.

Then there's time. Running experiments, teaching and writing papers and grants all consume more time than a 'normal' working week. It takes a very understanding partner to put up with that day after day. No one likes to be stood up for a romantic dinner because their partner had to run to the lab to check on gels.

Finally, there's stress, which tends to come from things that are beyond your control. Even if you crank out first-rate data, draft brilliant papers and write compelling grant applications, ultimate success often lies in someone else's hands. It is no mean feat keeping your composure in the face of such ceaseless challenges.

I have no solid data to support this thesis, other than a few anecdotes and some loose correlations. I'd love to be proved wrong, but need some better data. So please send any relevant stories or figures to me at naturejobseditor@naturedc.com.

Paul Smaglik, *Naturejobs* editor

CONTACTS
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:
Muriel Lestringuez (4994)
UK/Ireland/Italy/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:**
Tom Hancock
***Naturejobs* online production:**
Catherine Alexander

US Head Office, New York
75 Varick Street,
9th Floor,
New York,
NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

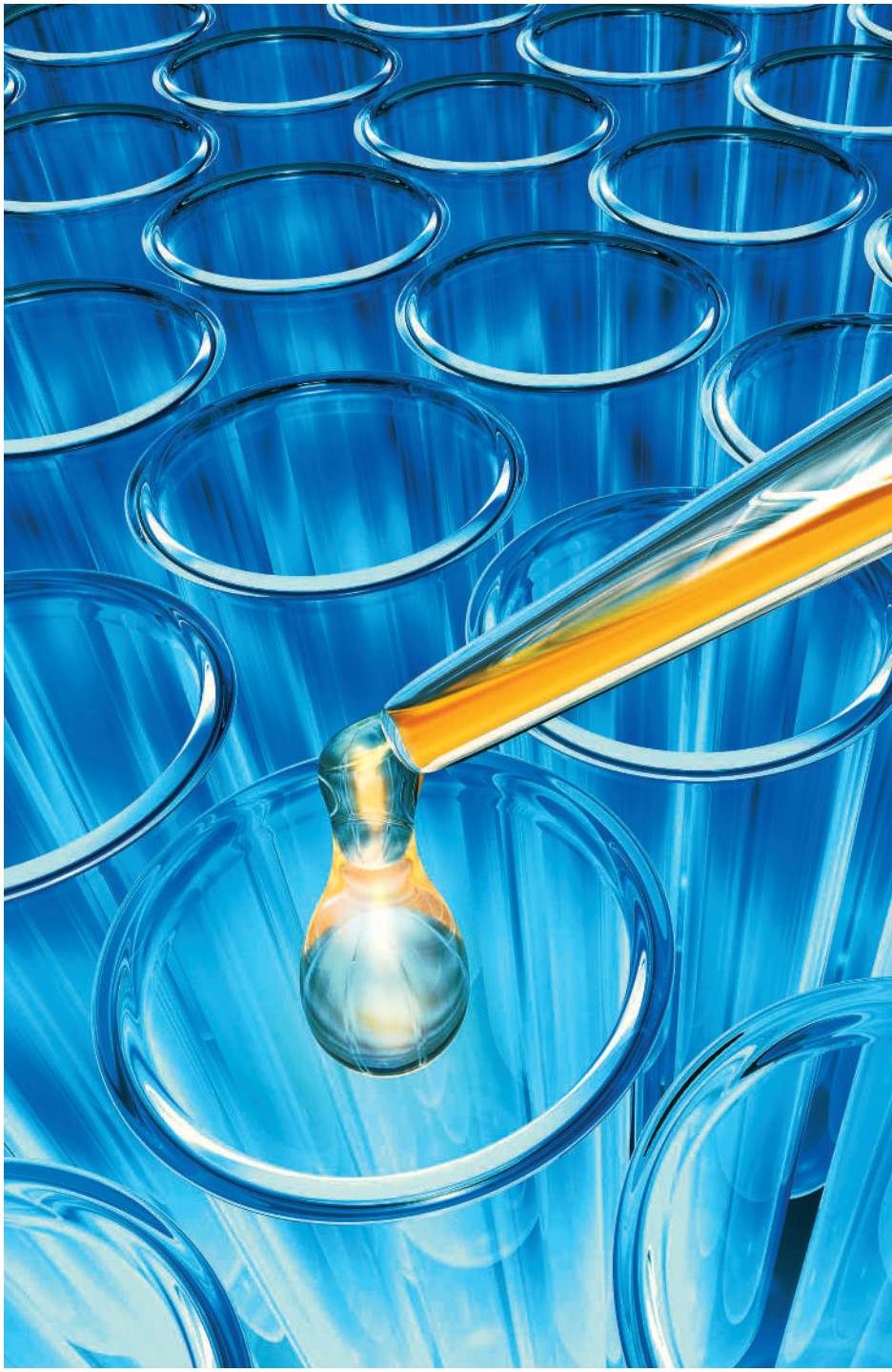
US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

CHEMISTRY IN CONTEXT

Rumours of the demise of chemistry are misplaced — the 'science of everything' prevails, if sometimes under new guises. **Ricki Lewis** reports.



M. KULKA/ZEFA/CORBIS

To the untutored eye, it may seem as though chemistry is past its prime — fading into the shadow of its siblings physics and biology, or subsumed into new disciplines such as chemical biology. Certainly, four of the past five Nobel prizes in the field have celebrated work with a decidedly cellular flavour. But instead of vanishing, chemistry is merely reinventing itself as its importance to other fields becomes increasingly recognized.

"Chemistry splits into thousands of things, and then some begin to have their own names. Whether it is or isn't chemistry depends on your point of view," says Arnold Thackray, president of the Chemical Heritage Foundation in Philadelphia. The branching of chemistry into many subdisciplines means that although there may be fewer traditional jobs in companies that focus on manufacturing compounds, there may be more opportunities in roles that straddle disciplines and combine skills. Flexibility, creativity and a constantly updated skill set are key to succeeding in these positions.

Combinatorial technique

Combining technical strengths can be valuable for positions that blend chemistry and biology. The trajectory for a chemical biologist, for example, is typically a PhD in organic chemistry with experience in synthesizing analogues of biomolecules, plus postdoctoral work using organic-chemistry approaches to biological problems.

"Today's chemical biologist might synthesize large libraries of small molecules and use them in biological screens as inhibitors," says Reid Gilmore, a biochemist at the University of Massachusetts Medical School in Worcester. "A chemical biologist pursues basic research, whereas a medicinal chemist synthesizes drugs." To make the transition from organic chemist to chemical biologist, he suggests taking biology and biochemistry courses. "Combined strength in organic chemistry and biochemistry is not that common," he notes.

But whatever strand of chemistry you chose to pursue, you will need more than just scientific savvy to succeed — especially if you are working in industry. "The three essential 'softer' skills are teamwork, flexibility and critical thinking," says Martha Collins, director of new applications in the Materials Research Center at Air Products and Chemicals in Allentown, Pennsylvania. "For an industrial chemist, it is now normal nowadays to change your main focus." Collins did her doctoral work on fibres, but has worked in coatings, surfactants, crosslinking and emulsion polymers, and now manages 50 researchers. To catalyse flexibility, Air Products' career-development programme gives its scientists stints in other areas, such as marketing. "You have to be able to take something that you are doing in the lab and see how it will be used, then design it the right way," Collins adds.

Strength in diversity

The applications to which skills in chemistry can be put are becoming increasingly diverse, from designing drugs to restoring antiques. Projects can vary even on a day-to-day basis. "A new chemist who wants to focus on one area for the rest of his or her life will have a hard time here," says Greg Chambers, global technology leader for polymer and chemical

CONSULTING ON CHEMISTRY

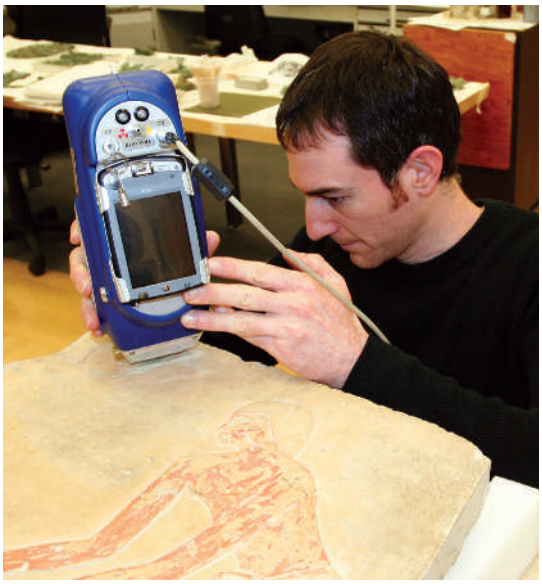
One option for chemists seeking an alternative career is to move into consultancy. "A consultant gives people a better understanding, at the atomic or molecular level, of their products and processes," says Joseph Sabol, who left academia and now runs a chemical-consulting business in Racine, Wisconsin. Consultants need to develop "a good seat-of-the-pants knowledge of a wide range of chemistry, physics, engineering, manufacturing and industrial processing", Sabol notes. For example, one of his clients was a shampoo designer run by non-scientists who needed a crash course in pH.

Opportunities can arise unexpectedly. Heather Harris's career as a forensics consultant was born out of a tragedy. At university in 1995, the professor who taught her physical-anthropology class was called to the site of the Oklahoma City bombings to analyse evidence. Harris was so inspired by the work that she went on to earn a master's in forensic science.

But she disliked working in a crime lab, so she went to law school. Like Sabol, she has to constantly seek out work, but finds it "when there is an issue related to chemical analysis of evidence", she says. Harris loves the challenge of "trying to find the little twist that could help the client, and putting together the pieces of a puzzle". **R.L.**

E. A. SOBCZAK

Joseph Sabol: enjoying consultancy.



AIR PRODUCTS AND CHEMICALS INC

J. ARSLANOGLU; EMORY UNIVERSITY

Chain reaction: (clockwise from top) Eric Breitung, Dennis Liotta and Heather Harris see a range of opportunities for chemists.

technologies at General Electric's Global Research Center in Schenectady, New York. "A chemist might work on silicones one day and organic light-emitting diodes the next, and have a great time doing it."

Even a company with a specific niche can offer eclectic opportunities. Cincinnati-based Procter & Gamble, for example, focuses on surface chemistry. But opportunities range from an analytical chemist to study the effects of colorants on hair to an organic chemist to work on lithium batteries.

For chemists looking to combine the excitement of discovery with the satisfaction of social responsibility, new business models may offer fresh opportunities compared with career paths at typical drug companies. For example, iThemba Pharmaceuticals, based in South Africa, was created as a joint venture between drug-discovery firm Chimerix of Durham, North Carolina, and universities in Britain and the United States to address public-health problems in sub-Saharan Africa. The collaboration will attempt to pursue treatments that big drug companies typically dismiss as unprofitable. In addition to researching small-molecule drugs, primarily for tuberculosis, HIV/AIDS and malaria, iThemba will refine manufacturing processes for generic drugs. Soon, the venture will be looking to hire several medicinal and synthetic chemists, says Dennis Liotta, its co-founder and professor of chemistry at Emory University in Atlanta, Georgia. "The goal is to be a financial success but also to make a positive difference to people's health," says Liotta. "It's a different kind of job. The pay will be competitive, but it will be a very noble task."

Art and science

Interdisciplinary approaches and niche industries are creating jobs for chemists in many places. When Eric Breitung, an inorganic chemist with industrial experience in coatings and thin films, had to relocate to New York City, he answered an advert from the Metropolitan Museum of Art. The museum was looking for a research fellow to solve a problem in the preservation of face-mounted photographs. Framing photos that are 2 metres by 2 metres in glass is difficult because it is too heavy. So an acrylic sheet is normally glued to the picture with silicone. Face-mounting improves depth and colour saturation, but if the acrylic becomes scratched or damaged, the repairs involve polishing out the damage, which can distort the image. Chemists such as Breitung are looking for other ways to make repairs without causing distortion.

Chemistry jobs increasingly comprise an array of specialized opportunities that allow practitioners to apply problem-solving skills. Yet what makes many chemists love their work is the joy of exploration and invention. "Every day is different," says Paul Docherty, a PhD student at the University of Oxford, UK. "I might be using similar techniques, but every compound has its own character and behaviour. I work within rules, and use my imagination to bend those rules, or find new ones. I'm doing stuff every day that no one else has done before." ■

Ricki Lewis is a freelance science writer in Schenectady, New York.

NATURE|Vol 446|22 March 2007

MOVERS

David Schimel, chief executive, National Ecological Observatory Network, Washington DC

1992–2006: Senior scientist, National Center for Atmospheric Research, Boulder, Colorado

David Schimel has a reputation among his colleagues for his quick-witted turns of phrase. "Our telescope points everywhere at once" is one example of his quirky sense of humour. Phrases such as this — an allusion to how the Hubble telescope compares with the National Ecological Observatory Network's (NEON's) planned array of numerous sensors — are now becoming part of the fabric of the fledgling network, headquartered in Washington DC. Last November, Schimel was appointed as chief executive of the network, which will see a set of sensors and facilities monitor ecosystems throughout the United States.

This might sound like a curious post for a former linguistics student, but Schimel's brush with ecology — when he enrolled in a course to prepare for backpacking and camping trips — began a romance with natural science. He went on to major in biology and later added a degree in mathematics, once he realized that a quantitative approach was important to his work. After two years at the Marine Biological Laboratory's Ecosystems Center in Woods Hole, Massachusetts, and a PhD on grasslands at Colorado State University, Schimel joined the National Center for Atmospheric Research (NCAR) in Boulder, Colorado, in 1992.

It was during an almost six-year hiatus from NCAR, at the Max Planck Society in Germany and the International Ecology Institute, that Schimel gained a deeper understanding of the culture of European ecology, which includes lots of collaborative projects involving networks of sites, a hallmark of NEON. "We will have to find a way to emphasize people's individual creativity and mix it with collaborative building of the network," says Schimel.

NEON is tentatively scheduled to be operational by 2013, and will process data from 20 stations throughout the United States. One of the instruments at its disposal, for example, will be a 10-metre tower to measure carbon dioxide flux, detect nitrogen oxides and record leaf wetness.

But NEON has been criticized by some as being poorly designed and not worth the expense (see *Nature* 444, 420–421; 2006). Schimel says the scientific community will assess NEON's usefulness by, for example, how well the project monitors land use, the presence of invasive species, and ecosystems' response to natural and human-induced climate changes. Despite the scepticism, Schimel remains excited about the project's incredible potential and is eager to listen to the science community to determine how best to prioritize NEON's activities. ■

Matthew Nestel

SCIENTISTS & SOCIETIES

Researchers without frontiers

Mobility is an important component for personal and career development — and is particularly useful in Europe. The Marie Curie Fellowships, set up by the European Commission, fund PhD and postdoc researchers to work in a country other than their home nation. To assist with this process, the independent Marie Curie Fellows Association (MCFA) came into existence ten years ago and is pushing for policies that would ease the movement of researchers from one country to another.

To make Europe a more attractive workplace for increasingly mobile researchers, the association needs input and support from mobile researchers, principal investigators, policy-makers, universities and industry. Part of its remit is to bring together these disparate parties and establish stronger contacts between them.

The MCFA provides feedback to European policy-makers through newsletters, by taking part in scientific, science-policy and science-related conferences and by publishing policy statements. For example, the MCFA strongly supported the 'European charter for researchers' and the code of conduct for their recruitment, two documents that were formally adopted by the commission as a recommendation in March 2005.

The charter aims to ensure that the relationship between researchers, employers and funding organizations contributes to the generation, transfer and sharing of knowledge, and to the career development of researchers. The code of conduct aims to make selection procedures fairer and more transparent and proposes judging merit using factors apart from publication output, such as teaching, supervision, teamwork, knowledge transfer, management and public awareness activities.

The experience gained by Marie Curie fellows during their stints in different countries will help them to play an important role in building the European Research Area, an initiative to provide a European 'common market' for research and innovation. But fellows face challenges — their mobility can prove disruptive to their families, for example. To ease such difficulties and to help the fellows realize their potential within Europe, the MCFA is doing its utmost to ensure that mobile researchers will be able to enjoy the best possible working conditions. ■

Vanessa Díaz is vice-chair and Guggi Kofod is vice-secretary-general of the Marie Curie Fellows Association.

POSTDOC JOURNAL

Beginner's luck

I'm writing this entry at the end of the fourth week of my first postdoc position. It is too soon for grand claims but, despite expectations to the contrary, it seems that being a postdoc isn't that bad. In fact, I'd almost go so far as to say that I'm really enjoying myself. What's going on?

For one thing, it's liberating to know that there is no degree riding on my performance; there aren't going to be any assignments or exams standing between me and the continued receipt of my salary. Postdoc positions present their own challenges, of course, but I like to think that it is still too early to be worried about such things.

My optimism could be due to the fact that so much in my life has changed. I've changed cities, changed labs, changed supervisors and, most importantly, changed research direction. I've already stumbled across what looks to be a worthwhile project in my new field, something both intellectually stimulating and maybe, just maybe, medically useful. And there's nothing quite like the promise of a manuscript being submitted for publication in the not-too-distant future to help make research life look rosy.

And then there's the very real possibility that this period of the job is like the beginning of any new relationship — all fun and no work. Here's hoping that it stays fun for a long time to come. ■

Peter Jordan is a first-year visiting fellow at the National Institute of Diabetes

The inside track from academia and industry

STAYING POWER

Carving out a successful career as an academic is hard, but there are steps you can take to ease your way.

Moray Campbell

The odds of progressing from being a PhD student to holding a faculty post are at least 100:1. For those who manage to make the move, their future success will hinge on them working on the right research question at the right time and in the right place. And on top of that, they should also be helping their employer's organization to evolve.

One person rarely meets these criteria indefinitely at a single institution. As a result, people need to be aware of their 'institutional shelf-lives' and should see their host institute as just a sub-network in the larger academic landscape in which individuals, ideas and resources flow freely. Fortunately, unlike most industries, there are several indicators that can be used to measure performance — from publications and grant income to teaching revenue and student feedback. And these can aid researchers in their quest for mobility.

Scientists in the early stages of their career rarely achieve success in all of the performance criteria, so the decision of an appointment panel for a post tends to be based on current research attainment and future potential.

The challenge of succeeding in a post is compounded by the current trend for consolidation in the academic sector. In several European countries, including Denmark and Britain, institutions are merging as if they were multinational companies. Such conditions make career strategy even more crucial — and there are some basic rules you should consider.

"For most people in academia, the overwhelming experience is rejection — most commonly of grants and manuscripts."

Faced by a charging grizzly bear, you have only to run faster than your neighbour

All selection committees focus on publications — as much for potential as for current achievement. So postdocs and graduate students need to finish and submit papers before they hit the job market, and should be proactive about the order of author names on their papers. But to stand out, you'll need something extra, such as being active in local, national and international organizations, having supervisory experience or a teaching role. And think carefully about the choice of referees you use on your CV or résumé. These must include your current principal investigator, but if you can offer external and international scientists, it will suggest that you have an interactive and outgoing nature.

Once in post, the same skills and mind-set will sustain you. In my case, I continued training undergraduate scientists and ensured that my graduate students received the best training I could facilitate, including training visits to the labs of more senior collaborators. Joining small projects together develops more cohesive bodies of work to attract larger grants and sustain publications. Exclusive focus on research can be a vulnerable position, and combining that with teaching and administrative duties will justify promotion. Alone, none of these steps is dramatic. But together they generate a broad portfolio.

Avoid dynasties

The trick is to pick up new skills and insights while moving through your PhD and postdoc positions. PhD training develops generic skills, such as how to handle pressure or how to write. Ideally, a PhD student should experience how to release their innate creativity. Postdoc work develops greater intellectual and research skills combined with the ability to supervise others. Repeating your principal investigator's ideas, or remaining in their sphere of influence too long, will nearly always make it harder for you to be recognized for your achievements, leading to stunted intellectual and career development, and frustration. Ironically, gaining independence can be made more fraught by some principal investigators' 'king-maker' tendencies.

Travel broadens the mind

Mobility stimulates novel approaches and insight, so recruiting committees tend to want candidates with diverse training backgrounds, and often mistrust 'home-grown' talent. The desire for international exposure, although it can be over-hyped, can also be beneficially exploited by following the money abroad, succeeding in a new country and then seeking out opportunities offered by repatriation schemes. Mobility doesn't finish with a faculty appointment; I have reached my institutional shelf-life here and, after about ten years in Britain, I am ready to return to the United States.

Sometimes you're ahead, sometimes behind, but in the end the race is with yourself

For most people in academia, the overwhelming experience is rejection — most commonly of grants and manuscripts. So you should examine your career motivations carefully. Find and use mentors locally, nationally and internationally to test your ideas. Remember to take on reviewers' feedback no matter how negative and provocative (if they didn't understand your grant, it may be because you didn't take time to write it clearly).

I am trying to follow my own advice. My goals now are to get my work published in some top journals, take up a position in the United States — and then plan my next move. ■

Moray Campbell is at the University of Birmingham, UK.